UNIVERSITAT POLITÈCNICA DE CATALUNYA

Programa de Doctorat:

AUTOMÀTICA, ROBÒTICA I VISIÓ

Tesi Doctoral

Demand Modeling for Water Networks Calibration and Leak Localization

Gerard Sanz Estapé

Director: Dr. Ramon Pérez

Gener de 2016

Abstract

The success in the application of any model-based methodology (e.g. design, control, supervision) highly depends on the availability of a well calibrated model. There is no best or unique solution for the calibration problem as the methodologies are developed depending on which parameters have to be calibrated and the final use of the model. The main objective in this thesis is to develop an adaptive water distribution network model which both calibrates its demands online and discerns between faults and system evolution. The calibration is focused on demands due to their daily variability and continuous evolution.

The singular value decomposition is a powerful tool for solving the optimization problem. Additionally, the deep understanding of this tool allows to redefine the demand model. A novel demand model is proposed, where each individual demand is defined as a combination of demand components. These demand components are calibrated demand multipliers that represent the behavior of nodes in a determined geographical zone. The membership of each nodal demand to every demand component is produced naturally through the analysis of the singular value decomposition of the sensitivity matrix. The same analysis is also used to define the location of sensors for the calibration.

The calibration in water distribution networks needs to be performed online due to the continuous evolution of demands. During the calibration process, background leakages or bursts can be unintentionally incorporated to the demand model and treated as a system evolution (change in demands). To solve that, a leak detection and localization approach to be coupled with the calibration methodology that identifies geographically distributed parameters is proposed. The approach consists in comparing the calibrated parameters with their historical values to assess if changes in these parameters are caused by a system evolution or by the effect of leakage. The geographical distribution allows to associate an unexpected behavior of the calibrated parameters (e.g. abrupt changes, trends, etc.) to a specific zone in the network.

The set of methods proposed are exemplified through an academic dummy network to help the reader completely understand their fundamentals. Furthermore, three real water distribution networks situated in Barcelona and Castelldefels are used to evaluate the performance of the whole method with real systems and real data. The good results obtained show the potential of the developed method and the viability of the real-time calibration and leak detection and localization processes.

Keywords: Water distribution networks, Calibration, Demands, Sampling Design, Leakage, Singular value decomposition, Real applications.

Resum

L'èxit en l'aplicació de qualsevol metodologia basada en models (p. ex. disseny, control, supervisió) depèn, en gran part, de la disponibilitat d'un model ben calibrat. No hi ha una solució única o global per aquest problema, ja que les metodologies es desenvolupen en funció de l'ús final del model. L'objectiu principal d'aquesta tesi és desenvolupar un model adaptatiu per xarxes de distribució d'aigua que calibri les seves demandes de forma online mentre distingeix entre fallades i evolució del sistema. La calibració es centra en les demandes degut a la seva variabilitat diària i a la seva evolució continua.

La descomposició en valors singulars és una eina molt potent per resoldre problemes d'optimització. Addicionalment, la comprensió detallada d'aquesta eina permet redefinir el model de demandes. Es proposa un model de demandes innovador, on cada demanda individual es defineix com una combinació de components de demanda. Aquests components de demanda són multiplicadors de demanda que han estat calibrats, i que representen el comportament dels nodes en una zona geogràfica determinada. La pertinença de cada demanda nodal a cadascun dels components de demanda es produeix de forma natural mitjanant l'anàlisi de la descomposició en valors singulars de la matriu de sensibilitat. El mateix anàlisi també s'utilitza per definir la localització dels sensors per la calibració.

La calibració en xarxes de distribució d'aigua s'ha de realitzar en línia, ja que les demandes evolucionen contínuament. Durant el procés de calibració, les fuites latents o espontànies poden ser incorporades involuntàriament al model de demanda, i ser tractades com una evolució del sistema (canvi en les demandes). Per solucionar-ho, es proposa un mètode de detecció i localització de fuites que s'acobla a la metodologia de calibració que identifica components de demanda geogràfics. El mètode proposat consisteix en comparar els paràmetres calibrats amb els seus valors històrics per valorar si els canvis en aquests paràmetres es deuen a una evolució del sistema, o a l'efecte de les fuites. La distribució ge-

ogràfica permet associar un comportament no esperat dels paràmetres calibrats (p. ex. canvis sobtats, tendències, etc.) a una zona específica de la xarxa.

El conjunt de mètodes proposats s'ha exemplificat mitjanant una xarxa acadèmica molt simple per ajudar al lector a entendre completament els seus fonaments. A més a més, s'han utilitzat tres xarxes de distribució d'aigua situades a Barcelona i Castelldefels per avaluar el funcionament del mètode complet amb sistemes i dades reals. Els bons resultats obtinguts mostren el potencial de la metodologia desenvolupada i la viabilitat de la calibració de demandes i detecció i localització de fuites en temps real.

Paraules clau: Xarxes de distribució d'aigua, Calibració, Demandes, Distribució de sensors, Fuites, Descomposició en valors singulars, Aplicacions reals.

Acknowledgments

This dissertation was carried out at the research Centre for Supervision, Safety and Automatic Control (CS2AC in catalan), and was partially supported by the project FP7-ICT-2012-318556 (EFFINET) of the European Commission, by the project DPI2013-48243-C2-1-R (ECOCIS) and DPI2014-58104-R (HARCRICS) of the Spanish Government, and by the Polytechnic University of Catalonia. The model of the real network and data from the SCADA system were provided by the Barcelona Water Company AGBAR.

I would like to thank all the members of the CS2AC group, as well as the ESAII department, for converting my daily routine into a pleasant activity. I will always remember the breakfasts, "xocomàtiques", and Sant Tomàs excursions.

I am indebted to my supervisor Ramon Pérez for his involvement in my thesis, always keeping the same enthusiasm as the first day. Thank you for guiding me, helping me, and sometimes worrying about me more than I do. I could not be more proud to have been your first PhD student.

During my PhD, I had the opportunity to join the Centre for Water Systems (CWS) at the University of Exeter. I thank Dragan Savic and Zoran Kapelan for helping me out with my research, and to all people of the CWS for making me feel like I was at home. I would also like to thank Olivier Piller and Enrico Creaco for reviewing this work. Your comments have helped to improve the quality of the thesis.

Lastly, I also want to thank Sílvia and my family for their support, patience and encouragement during these years. Without them, this thesis would not have been a real fulfillment.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction and motivation

Since the beginning of times, nature has been evolving until it has reached an equilibrium state that preserves its sustainability. We, the humans, evolve so fast that we cannot mimic the nature sustainable evolution paradigm. For the human race evolution to be sustainable, we need to take the most from our capacity of generating knowledge, of deeply understanding our own existence. Therefore, we need models of what is happening around us and what are we doing. This can be applied to every system. One of them, which in fact affects both the nature and the human race, is water. "Water is everywhere. Water is the touchstone. Water is the linchpin of human health and ecosystem health" *Karen Bakker*. In this thesis I am going to focus on a small part of the linchpin: Water distribution networks.

Water distribution networks (WDN) are complex systems whose main purpose is to supply the system's users with the amount of water demanded under adequate pressure and quality conditions. These conditions have to be satisfied independently of the existence of faults in the network, such as leaks, bad water quality, or fraudulent consumptions.

Research in WDN covers a wide range of topics: WDN modeling, network optimization, background leakage modeling and control, smart demand metering, assessment, forecasting and management, asset management and performance modeling, real time monitoring, modeling and control, network vulnerability, reliability, resilience and risk analysis, leakage and energy management, transient analysis, water quality, contaminant intrusion and water security, network operation and maintenance, etc. All solutions provided in these fields make use of

models. Besides, the daily routine in water companies also takes advantage of hydraulic models to perform the control and management of their networks.

When talking about any kind of models, both researchers and practitioners agree that a good calibration of the models is required to obtain reliable solutions when using them. Water distribution network modeling has been one of the main focuses of research during the last decades, as it will be seen in Chapter 2. The question is... How did I end up doing research on WDN calibration?

Five years ago, in 2010, I was about to start my last year in the Automatic and Control Engineering degree. It was then when my (current) thesis supervisor, Ramon Pérez, suggested me to become part of the Advanced Control Systems group taking profit of a research scholarship. By then I had never seen a water distribution network model, nor launched an EPANET simulation. So I started working with hydraulic models, MATLAB, EPANET, sensors, demands, nodes, flows and heads, leaks, etc., and learning all I could about WDN.

During that year, I took part in the Real Time Network Monitoring project (AM901) of R+i Alliance, coordinated by CETAQUA, a Water Technological Centre. My main function was to implement the ideas of the research group, mainly sensor placement and leak detection methodologies. Results with simulated data were incredible, until we started playing with the model demands. We realized that the demand model had a huge impact on the leak detection and localization results, so I started analyzing these effects. From these studies arose my first research contributions (Pérez et al. [2011a], Sanz and Pérez [2012]) and, what is the most important, the motivation to start my own research.

As mentioned before, a lot of research has been done in WDN demand calibration. However, few of the reviewed literature takes profit of the demand calibration to detect events (mostly leaks) that occur in the WDN and may be masked by the calibration, and none of them do it in an online procedure, as proposed in this thesis. You cannot calibrate without considering the possible leaks in the network, nor try to detect and locate them without a good hydraulic model. We have tried to convert these tradeoffs into synergies: to calibrate the hydraulic model considering the possibility of leakage, which allows not only to obtain a good model, but also to be aware of faults in the network.

The reader will find in this thesis a novel way of defining demands depending on their geographical location and pressure/flow sensitivities. Furthermore, the thesis is intended to work as an educational book, presenting at each chapter an exemplification of the theoretical methodology using a dummy water distribution network. Nevertheless, the thorough theoretical development does not blur

the main objective of the thesis: the application of the methodology in real networks. In Chapter 6 the reader will find three examples of the application of the developed methodologies in real scenarios.

1.2 Main objectives

As it was pointed out in the previous section, the main goal of this thesis is to develop an adaptive on-line model for demands in WDN. This model will calibrate its parameters on-line, and will be able to discern between system evolution (e.g. changes in demands) and faults (e.g. leakages, obstructions). This type of model may be used for water management and prognosis. The secondary objectives involved in the generation of the adaptive model are:

- Parameterization: Determine the parameters that will be calibrated that allow to improve the hydraulic model while permitting to detect and locate anomalies in the network.

- Sampling design: Determine the type and location of sensors to guarantee the identifiability of the reparametrized system.

- Network skeletonization: Develop a process that reduces the model's elements while keeping the network behavior unchanged, according to the use of the model.

- Network calibration: Once the system is identifiable (thanks to the parameterization and sampling design), develop a methodology to calibrate online the network parameters.

- Fault versus parameters' evolution identification: Develop a fault detection method that analyses the calibration results to evaluate if changes in parameters are provoked by a fault in the network (mainly leakage), or by a system evolution (change in demands).

- Software development: An open source software package combining the EPANET water simulation engine with MATLAB will be programmed. The package will include all functionalities developed during the doctoral thesis.

- Methodology versatility: Analyze the applicability of the developed methodologies in multiple WDN.

1.3 Outline

The manuscript is organized in seven chapters. Chapter 2 presents the background theory and state of the art. Chapter 3 proposes a methodology to select the network parameters and sensors, which is used to redefine the calibration problem in Chapter 4. Next, Chapter 5 presents an approach to analyze the calibration results for leak detection and location. The complete methodology is applied to three different networks in Chapter 6. Finally, conclusions are drawn in Chapter 7. The outline of this thesis is briefly detailed next:

- Chapter 2: This chapter presents a thorough review of the state of the art in water distribution network calibration and leakage detection and localization. Background theory on WDN modeling is also presented. All the content is mainly focused on water demands.

- Chapter 3: An approach based on the singular value decomposition theory is proposed to use the information of the system sensitivity matrix to define the parameters that will be calibrated, and the type and location of sensors that are needed to guarantee the system identifiability.

- Chapter 4: The parameterization and sampling design results obtained in Chapter 3 are used in this chapter to redefine the calibration problem. The singular value decomposition is also used in this chapter to solve the general inverse problem. An online calibration structure is proposed.

- Chapter 5: This chapter presents an approach to analyze the calibrated parameters in order to detect and locate leakages in the water distribution network, or other faults that cause similar abnormal behaviors. The procedure compares the calibrated components values with their historical values to asses if changes are due to demand evolution or to faults.

- Chapter 6: The methodologies developed are applied to three real networks located in Barcelona and Castelldefels. Real data are used in the three scenarios to apply the precalibration and calibration processes (in the three networks), and the leak localization process (applied to the only network where a known leak scenario was available).

- Chapter 7: The main contributions and conclusions of the thesis are summarized, taking into account the future work and questions to be answered found in the literature review.

1.4 Related publications

The author has published the research related with the thesis in multiple conferences and journals. A list of published works that influenced or derived from the thesis development are presented next:

- Pérez et al. [2011a]: Conference paper that analyzes the effect of the demand model on the leak detection and localization accuracy. This work motivated the thesis development.
- Sanz and Pérez [2013a]: Website with the last updated version of the software developed by the author.
- Sanz and Pérez [2013b]: Conference paper presenting a first version of the developed software applied to a case study for pressure control.
- Sanz and Pérez [2014a]: Conference paper that presents a first version of the calibration methodology. In this paper the calibration of nodal demands is compared to the calibration of user-based demand patterns. Chapter 4.
- Sanz and Pérez [2014b]: Conference paper presenting the parameterization process with binary memberships, and the sampling design methodology applied for a single working point. Chapter 3.
- Sanz and Pérez [2014c]: Conference paper that compares the calibration results using geographical demand components with binary memberships to the calibration results using user-based demand patterns. Chapter 3 and Chapter 4.
- Pérez and Sanz [2014]: Conference paper that compares the sampling design methodology presented in this thesis with another methodology based on leak detection. Chapter 3.
- Sanz and Pérez [2015a]: Journal paper including the parameterization process considering binary, positive and free memberships, and the application to a real network with synthetic data. Chapter 3 and Chapter 4.
- Sanz et al. [2015]: Journal paper that presents the leak detection methodology based on the analysis of online calibrated demand components. Chapter 3, Chapter 4 and Chapter 5.
- Sanz and Pérez [2015b]: Conference paper that analyzes the demand component calibration results depending on the types of sensors used. Chapter 3 and Chapter 4.

- Pérez et al. [2015]: Conference paper that analyzes the uncertainty from real sensors' measurements and replicates the computed uncertainty using demand components.
- Sanz and Pérez [2015-SUBMITTED]: Journal paper that reviews the demand components calibration online methodology and deeply analyzes the application to a real scenario. Chapter 3, Chapter 4, and Chapter 6.

Chapter 2

Review of the state of the art

This chapter presents a review of the state of the art in all topics treated in the thesis. The existing types of water networks models and the modeling of their behavior are described, focusing on steady state models, which have been considered in this thesis. Next, the global hydraulic calibration problem is presented and the existing solutions reviewed. The supporting techniques required by the calibration methods such as sensitivity matrix calculation, network reduction, sampling design and uncertainty calculation, are also presented, and their current state of the art reviewed. Finally, a quick review to fault detection (leakage) references is presented.

2.1 Water distribution network models

Models allow us to simulate the behavior of WDN for many purposes: design of water infrastructures, predictive control and optimization, or state monitoring to predict, detect and locate possible faults, from both hydraulic (leaks) or quality (chlorine level) natures. The type of model used depends on its final use.

Water distribution network models can be classified depending on the dynamics involved: dynamic models and static models. Some WDN models characterize the transients in pipes, valves and pumps (Liggett and Chen [1994]). The analysis of transients is used to know the dynamic behavior of the network. Dynamic models have been used for multiple objectives: Vítkovský et al. [2000] used these models for calibration of roughness, whereas Kapelan et al. [2003a] used inverse transient models for leakage detection. The main drawbacks of dynamic models are the need of huge amounts of data, and the high computational power required.

Slow transient models (Shimada [1992], Piller and Propato [2006]) or rigid

water column models are important when there is mass oscillation in the system still being incompressible (slow valve closures, tanks located in near locations, pumps closures between two resource nodes, etc.). These oscillations may occur in pseudo real time modeling with observation time lower than 5 minutes.

When the number of pipes, pumps and valves increases, the network tends to become steadier and the transients lose importance. Furthermore, the sensors' sampling times use to be much lower than the network elements' dynamics, as WDN are always pressurized. Computer-based supervision and control applications in huge networks assume that network behavior is described by steady state models concatenated in an extended period simulation (EPS) (Brdys and Ulanicki [1994], Walski et al. [2003]).

Steady state models are the most used in water companies for design, supervision and control. In this thesis, the calibration effort is focused on these models. Steady state models are formed by a set of interconnected elements which represent the real network components. The simulation of these elements allows to obtain an accurate estimation of variables such as pressure, flows, chlorine concentration, etc., if the model is well calibrated. An exhaustive description of all the elements can be found in the EPANET users guide (Rossman [2000]), a fully equipped, extended period hydraulic analysis package that has been used in this thesis.

2.1.1 Hydraulic equations and matrix model

The governing laws for flow in WDN under steady conditions are *conservation of mass* and *energy*. The *law of mass conservation* states that the rate of storage in a system is equal to the difference between the system's inflow and outflow. In pressurized WDNs, no storage can occur within the pipe network, although tank storage may vary over time. Therefore, in a pipe or a junction node, the inflow and outflow must be balanced. For a junction node,

$$\sum_{j=1}^{n_{pip}} B_{ij} \cdot \mathbf{q}_j = \mathbf{d}_i, \tag{2.1}$$

where n_{pip} is the number of pipes in the network, $\mathbf{q}_j$ is the flow of pipe j, and $\mathbf{d}_i$ is the consumption of node i. The coefficient $\mathbf{B}_{ij}$ indicates if pipe j is connected to node i:

- 1 if flow of pipe j enters node i.
- 0 if pipe j and node i are not connected.

- -1 if flow of pipe j leaves node i.

These coefficients generate the incidence matrix $\mathbf{B}$, which defines the connections between nodes and pipes.

The *energy conservation law* states that the difference in energy between two nodes is equal to the energy added to the flow in the components between these points minus the frictional and minor losses. The relationship between pipe flow and energy loss caused by friction in individual pipes can be computed using one of three formulas: Hazen-Williams, Darcy-Weisbach, and Chezy-Manning. The general relationship is of the form:

$$\Delta \mathbf{h}_{ij} = \mathbf{h}_i - \mathbf{h}_j = R_{ij} \cdot \mathbf{q}_{ij} \cdot |\mathbf{q}_{ij}|^{r-1}, \tag{2.2}$$

where $\Delta \mathbf{h}_{ij}$ is the headloss in pipe connecting nodes i and j, R_{ij} is the resistance coefficient that depends on the pipe's diameter, length and roughness, $\mathbf{q}_{ij}$ is the pipe flow rate, and r is the flow exponent. Expressions for the resistance coefficient and values for the flow exponent for each of the mentioned formulas are listed in Rossman [2000]. The energy balance for any path can be expressed as,

$$\sum \Delta \mathbf{h}_{ij} = \delta \mathbf{E}_p, \tag{2.3}$$

where $\delta \mathbf{E}_p$ denotes the energy difference between the starting and final nodes of the path p. The summation is carried out over all links of the path. For loops, $\delta \mathbf{E}_p = 0$, as the starting and ending node is the same.

The equations of the water distribution network can be defined in matrix form (Brdys and Ulanicki [1994]). The flow continuity law defined in Equation 2.1 can be represented in matrix form using the incidence matrix for junctions $\mathbf{B_j}$:

$$\mathbf{B_j} \cdot \mathbf{q} = \mathbf{d} \tag{2.4}$$

The formula for tanks can be found in Brdys and Ulanicki [1994], although it will not be used in this thesis as the available WDN do not include tanks.

On the other hand, the head-flow relationship (Equation 2.2) for a pipe can be written, according to the Hazen-Williams formula:

$$\mathbf{h}_i - \mathbf{h}_j = \frac{1.21216 \cdot 10^{10} \cdot L}{C^{1.852} \cdot D^{4.87}} \cdot \mathbf{q}_{ij} \cdot |\mathbf{q}_{ij}|^{0.852} \tag{2.5}$$

where L is the length in m and D is the diameter in mm of the pipe connecting both nodes; and C is the pipe roughness coefficient. If the head units are meters,

then the flow is in l/s. The flow can be isolated:

$$\mathbf{q}_{ij} = G_{ij}^{0.54}(\mathbf{h}_i - \mathbf{h}_j)|\mathbf{h}_i - \mathbf{h}_j|^{-0.46} \tag{2.6}$$

where G_{ij} is the pipe conductivity, calculated as:

$$G_{ij} = \frac{1}{R_{ij}} = \frac{C^{1.852} \cdot D^{4.87}}{1.21216 \cdot 10^{10} \cdot L} \tag{2.7}$$

A diagonal matrix $\mathbf{C}$ containing in its diagonal elements the nonlinear parameter of each pipe can be defined as:

$$\mathbf{C}_{pp} = G_{ij}^{0.54}|\mathbf{h}_i - \mathbf{h}_j|^{-0.46} \tag{2.8}$$

where $\mathbf{C}_{pp}$ is the nonlinear coefficient for pipe p, which connects node i with node j. By means of this matrix, the head-flow equation can be transcribed in matrix form:

$$\mathbf{q} = -\mathbf{C} \cdot (\mathbf{B_j}^T \cdot \mathbf{h} - \mathbf{B_f}^T \cdot \mathbf{h_f}) \tag{2.9}$$

where $\mathbf{C}$ is a diagonal matrix with the C_{pp} coefficients, and $\mathbf{B_f}$ is incidence matrix for fixed head nodes .

The matrix form of the equations governing the water distribution system can be obtained by joining Equation 2.4 and Equation 2.9:

$$-\mathbf{B_j} \cdot \mathbf{C} \cdot (\mathbf{B_j}^T \cdot \mathbf{h} - \mathbf{B_f}^T \cdot \mathbf{h_f}) = \mathbf{d} \tag{2.10}$$

2.1.2 Demand models

Once the WDN model is available, a demand model has to be defined. Nodes in WDN models represent an aggregation of multiple users. Each of these users may be of different type, e.g. domestic, commercial, etc. Users of the same type are usually assumed to consume water in the same way, following a predetermined diurnal demand pattern. The consumption of each user is computed by multiplying the pattern coefficients with the user's base demand, i.e. the user's average water consumption, computed from billing information. Once this is done, demands that are associated with a certain network node are aggregated, resulting in the total nodal consumption at a given point in time. To simplify, the demand at a

network node, assuming a single user per node is computed as

$$\mathbf{d}_i(t) = \frac{bd_i}{\sum_{j=1}^{n_d} bd_j} \cdot \mathbf{p}_{a \to i}(t) \cdot \mathbf{q_{in}}(t) \tag{2.11}$$

where bd_i is the base demand of node i, n_d is the number of nodes in the network, $\mathbf{p}_{a \to i}(t)$ is the value of diurnal pattern a associated with user i at time t, and $\mathbf{q_{in}}(t)$ is the total network consumption metered at sample t.

However, the information on different types of users associated with a given network node and their diurnal patterns is not always available in practice. Quite often, the only information available is the consumption aggregated during a period of time (usually monthly or quarterly). This low temporal resolution information on demands can still be used to compute the base demand of each user. The base demand of a node is computed from the sum of the base demands of consumers aggregated in this node. The basic model presented in Equation 2.12 uses the nodal base demands, together with the total network consumption metered at the network inputs, to calculate the instantaneous demand of each node at each sample.

$$\mathbf{d}_i(t) = \frac{bd_i}{\sum_{j=1}^{n_d} bd_j} \cdot \mathbf{q_{in}}(t), \tag{2.12}$$

The approach presented in Equation 2.12 considers that all demands have the same behaviour, which is determined by $\mathbf{q_{in}}$. Besides, the basic model cannot explain the daily variation of the relative pressure behavior between two areas in the network. This thesis proposes a new approach to model nodal demands depending on their geographical location by means of the calibration of demand components. A detailed explanation is presented in Chapter 3.

2.1.3 Simulation

Steady state simulations allow to determine the operating behaviour of a system under static conditions: fixed set of reservoir levels, tank levels, and water demands. Steady state simulations can be concatenated during time to generate extended-period simulations, where the only dynamics involved are the ones coming from tanks filling or emptying.

These simulations require an iterative technique to solve the non-linear equations involved. A key property of the nodal model is that it possesses a solution and the solution is unique (Brdys and Ulanicki [1994]). For existence of the solution it is sufficient to have at least one fixed-head node in each set of connected

elements (then the rank of $\mathbf{B_j}$ will be maximum and equal to the number of junction nodes). The most commonly used approach to solve the static simulation is the application of the Newton-Raphson algorithm.

Many software are currently available to model, simulate and control WDN:

- Piccolo (Safege) is a flexible WDN modelling and simulating software that allows the users to easily build their networks. It allows to diagnose structural weaknesses of the WDN, to study the impact of management policies on the residence time and the quality of water supplied to the final consumer, or to control the network performance, among others. Piccolo uses the hydraulic algorithm from Carpentier et al. [1987].
- EPANET (EPA) is a free software that models water distribution piping systems. The main advantage of EPANET is that it is a public domain software that can be freely copied and distributed. EPANET provides a user interface that allows to build networks, perform extended period simulation, model constant or variable speed pumps, etc. EPANET offers the EPANET Programmer's Toolkit, a dynamic link library (DLL) of functions that allows developers to customize EPANET to their own needs. The functions can be incorporated into Windows applications written in C/C++, Delphi, Pascal, Visual Basic or any other language that can call functions within a Windows DLL. EPANET solver is based on the gradient algorithm (Todini and Pilati [1988]).
- WaterCAD (Bentley) is a robust and easy-to-use water distribution modeling program that can be customized with additional modeling platforms. It can run in windows stand-alone platform or within other environments (AutoCAD, ArcGIS, SCADA). WaterCAD offers from fire flow, flushing, criticality, and constituent concentration analyses, to energy cost management and pump modeling. WaterCAD uses an improved version of the EPANET engine.

Additional water distribution system simulators are available but have not been included in this quick review. The election of one or another is based on the users' preferences, needs and budget. In this thesis EPANET software has been used to perform the WDN simulations.

All the presented software have limitations when programming in their own environment. However, the EPANET Programmer's Toolkit offers the possibility to simulate the model from an external software. One of the most used software in control engineering to develop and validate methodologies is Matlab, a high-level language and interactive environment for numerical computation, visualization,

and programming, used in a range of applications, including signal processing and communications, image and video processing, control systems, test and measurement, computational finance, etc. Furthermore, it includes tools for building applications with custom graphical interfaces.

A modular software combining the EPANET simulations with the computational power and the versatility of Matlab has been developed by the author. The software is available in http://cs2ac.upc.edu/en/training-benchmarks/simulador-de-xarxes-de-distribucio-daigues. The objectives of sharing the software are:

- To encourage the use of network models for analysis, supervision, control, design, etc., by companies, universities and research centers.
- To present the work done by CS2AC research group in this topic.
- To help solving specific problems of the water industry.

The water distribution network simulator format is:

- Main graphical interface: This folder contains the main functions of the simulator. It allows getting data from the model and its parameters, as well as perform simple simulations.
- Pressure control module: This extra module allows performing monovariable and multivariable pressure control over the water distribution network.
- Calibration module: This extra module allows calibrating the demand patterns of a water distribution network from a set of measurements.
- Demo network GNet2.
- Demo network GNet3.
- Quick reference guide of the simulator and its modules.
- Communication toolkit EPANET-Matlab.

A set of four practical sessions is available in order to introduce the user in the use of the interface:

- Session 1: EPANET simulator. Requires the download of the EPANET software http://www2.epa.gov/water-research/epanet.
- Session 2: Simulation and analysis module.
- Session 3: Pressure control module.

- Session 4: Demand calibration module

The software has been used in engineering degree and MSc courses to introduce WDN to the students, while reinforcing general modeling and control knowledge, and in a European capacity building project (MoICT). More information about the software can be found in Sanz and Pérez [2013a] and Sanz and Pérez [2013b].

2.1.4 Network reduction

Generally, water network models are automatically generated from Geographic Information System (GIS). This direct translation generates a model with a huge number of elements which do not have any impact on the network behavior. The main aim of a reduced model is to preserve the non-linear hydraulic behavior of the original network and approximate its operation accurately under different conditions. There are different methods for reducing the complexity of the model, such as skeletonization, decomposition, usage of artificial neural networks (ANNs) meta-model, or variables elimination.

Skeletonization is the process of selecting for inclusion in the model only the parts of the hydraulic network that have a significant impact on the behavior of the system (Walski et al. [2003]). The level of skeletonization depends on the intended use of the model. The reduced models have been called "surrogate networks" or "grey boxes" (Shamir and Howard [1977]). Eggener and Polwoski [1976] did the first study of skeletonization when they systematically removed pipes from a model to test the sensitivity of results. Brandon [1984] suggested three heuristic rules that can be used to carry out the skeletonization process: (1) relatively small demands along any pipe are added to the node at the end of the pipe; (2) pipes with small diameters are eliminated, and the area that is fed by them is represented by a single node; and (3) a group of adjacent nodes with similar pressures is reduced to one node. Hamberg and Shamir [1988] proposed an approach for reducing the size of the models for the preliminary design phase based in a step-wise combination of the system elements. Saldarriaga et al. [2009] skeletonized the network using the resilience concept. Walski et al. [2003] proposed an automated skeletonization process consisting in:

- Removing simple pipes: Pipes are removed from the system based on size or other criteria without any consideration of their effects on demand loading or hydraulic capacity.

- Removing branch pipes: Dead-end branches not containing tanks are trimmed back to a node that is part of a loop. This type of removal has no effect on the carrying capacity of the remainder of the system.
- Removing pipes in series: Pipes connected in series are replaced by an equivalent pipe which produces the same head-loss. Removed nodes split their demands between the two nodes at the ends of the resulting pipe. A cut-off may be considered in order not to remove nodes with large demands.
- Removing parallel pipes: As in the previous case, an equivalent pipe replaces the parallel ones. New pipe's parameters have to be calculated. No effect on demands is produced in this process.
- Removing pipes to break loops: Pipes with the lowest carrying capacity are removed for breaking loops. This action produces a loss of the system capacity. This process is performed to obtain a branch system that can be further skeletonized.

Non-pipe elements can also be removed but with some considerations (Walski et al. [2003]). Using these basic steps, automated skeletonization reduces the network until a stopping criteria defined by the user is achieved. This stopping criteria is chosen depending on the use of the model.

Swamee and Sharma [1990] proposed a simplification of the network by decomposing it in subsystems with one input in order to reduce the computational cost on the design of the water distribution system (WDS). Deuerlein [2008] introduced the network reduction process as a decomposition of the network graph according to its connectivity properties.

Anderson and Al-Jamal [1995] presented a parameter fitting approach. They reduced the network by calculating two parameters' vectors representing the nodal demands and the links conductances. An objective function was formulated for maximizing the accuracy of the simplified network. Rao and Alvarruiz [2007] captured the domain knowledge of hydraulic simulation model using ANNs for predicting the consequences of different control settings on the performance of the WDS. Broad et al. [2010] presented a systematic methodology using metamodels and ANNs. The purpose of the metamodels is not to approximate the entire simulation model, but to obtain a relationship among variables that contribute to the fitness (e.g. energy consumption).

Variable elimination is based on a mathematical formalism. Some of the system variables can be eliminated from the system of non-linear differential equa-

tions that represent the mathematical model. Martinez et al. [2014] presented an extended version of Ulanicki et al. [1996], proposing an algorithm involving linearisation, Gaussian elimination and a reconstruction of a reduced non-linear model. Paluszczyszyn et al. [2011] presented an implementation of the latter algorithm for integration of the model reduction module with an on-line optimization strategy.

Appendix A presents the method used in this thesis to reduce networks with large number of pipes and nodes, based on the skeletonization in Walski et al. [2003].

2.2 Sensitivity calculation

Sensitivity coefficients, the partial derivatives of head and flow with respect to each of the parameters, play an important role in the solution of the inverse problem (Yeh [1986]). Besides, the methodologies developed in this thesis require the calculation of the sensitivity matrix.

Some of the existing general methods for the calculation of the sensitivity matrix are (Kapelan et al. [2003b]): (1) Influence coefficient method (or influence coefficient method or perturbation method), (2) sensitivity equation method, (3) variational method (or adjoint method), and (4) automatic differentiation method.

The influence coefficient method uses the concept of parameter perturbation. At each simulation, one of the model parameters is perturbed (Bard [1974]), and the outputs measured. This method can be easily implemented, though computationally slow and relatively inaccurate when compared to other methods. $N+1$ simulations are required, where N is the number of parameters in the model.

In the sensitivity equation method, a set of sensitivity equations are obtained by taking the partial derivatives with respect to each parameter in the governing equation and initial and boundary conditions, when using slow transient models. For the steady state model there is no need to solve a sensitivity equation, as there exists an explicit formula, which requires a solution of the forward problem (heads and flows) prior to the determination of unknown sensitivities. The calculated sensitivities are very accurate (Kapelan et al. [2003b]).

The adjoint method computes relevant sensitivities once Lagrance multipliers are determined from a set of adjoint equations, which are derived from the basic WDS hydraulic model equations in subsection 2.1.1. This method also has a

high accuracy, and requires a solution of the forward problem prior to the determination of unknown sensitivities. The number of additional simulations is the number of Lagrange multipliers or adjoint moments, which is equal to the number of rows in the governing equation. The sensitivities are calculated only for selected (not all) model predicted variables.

The automatic differentiation method (Griewank et al. [1998]) is based on the differentiation of algorithms. Despite the good accuracy and computational performance, it produces a lengthy and complex computer code and requires a large number of changes to the source code of the appropriate hydraulic model (Chen [1995]).

Finally, a matrix analysis of the WDN linearized model where only one simulation is required at each iteration is proposed in Cheng and He [2011]. This latter approach is used to compute the sensitivity matrix in this work, with the improvement provided by Piller [1995], but alternative techniques can be applied.

2.2.1 Equation perturbation

The method proposed by Cheng and He [2011] consists in the perturbation of the linearized model. The matrix model of the water distribution network is defined by Equation 2.10. Considering an error $\Delta\mathbf{d}$ in predicted demands $\mathbf{d}^p$ that produces an error $\Delta\mathbf{h}$ in predicted heads $\mathbf{h}^p$, the linear approximation is computed as

$$\begin{aligned} -\mathbf{B}\mathbf{C}\mathbf{B}^T(\mathbf{h}^p + \Delta\mathbf{h}) &= \mathbf{d}^p + \Delta\mathbf{d} \\ -\mathbf{B}\mathbf{C}\mathbf{B}^T\Delta\mathbf{h} &= \Delta\mathbf{d} \end{aligned} \tag{2.13}$$

Defining $\mathbf{A} = -\mathbf{B}\mathbf{C}\mathbf{B}^T$,

$$\Delta\mathbf{h} = \mathbf{A}^{-1}\Delta\mathbf{d} \tag{2.14}$$

Matrix $\mathbf{A}^{-1}$ in Equation 2.14 is the complete sensitivity matrix relating changes in demands to changes in heads. If only a few head measurements are available, the new system of equations is redefined in Equation 2.15.

$$\mathbf{A}_{\mathbf{mh}}\Delta\mathbf{d} = \Delta\mathbf{h}_{\mathbf{mh}} \tag{2.15}$$

Matrix $\mathbf{A}_{\mathbf{mh}}$ and vector $\Delta\mathbf{h}_{\mathbf{mh}}$ are extracted from matrix $\mathbf{A}^{-1}$ and vector $\Delta\mathbf{h}$, respectively. Subscript $_{\mathbf{mh}}$ refers to measured heads. Matrix $\mathbf{A}_{\mathbf{mh}}$ is the sensitivity matrix relating changes in demands to changes in measured heads.

The same process is applied to the flow equation defined in Equation 2.9.

Considering an error $\Delta\mathbf{q}$ in predicted flows $\mathbf{q}^p$ that produces an error $\Delta\mathbf{h}$ in predicted heads $\mathbf{h}^p$, the perturbation equation is computed as

$$\begin{aligned} \mathbf{q}^p + \Delta\mathbf{q} &= -\mathbf{C}\mathbf{B}^T(\mathbf{h}^p + \Delta\mathbf{h}) \\ \Delta\mathbf{q} &= -\mathbf{C}\mathbf{B}^T\Delta\mathbf{h} \end{aligned} \tag{2.16}$$

Replacing $\Delta\mathbf{h}$ with the one defined in Equation 2.14:

$$\Delta\mathbf{q} = -\mathbf{C}\mathbf{B}^T\mathbf{A}^{-1}\Delta\mathbf{d} \tag{2.17}$$

Matrix $\mathbf{A_{mq}}$ and vector $\Delta\mathbf{q_{mq}}$ are extracted from the result of $-\mathbf{C}\mathbf{B}^T\mathbf{A}^{-1}$ and vector $\Delta\mathbf{q}$, respectively. Subscript $_{\mathbf{mq}}$ refers to measured flows. Matrix $\mathbf{A_{mq}}$ is the sensitivity matrix relating changes in demands to changes in measured flows.

$$\mathbf{A_{mq}}\Delta\mathbf{d} = \Delta\mathbf{q_{mq}} \tag{2.18}$$

Both head/demand and flow/demand sensitivity matrices are computed using demand allocation based on quarterly billing, as described in Equation 2.12.

Piller [1995] showed by the implicit theorem that the sensitivity of head with respect to demand is the matrix:

$$\Delta\mathbf{h} = (-\mathbf{B}\tilde{\mathbf{C}}\mathbf{B}^T)^{-1}\Delta\mathbf{d} \tag{2.19}$$

where matrix $\mathbf{C}$ is also perturbed. The diagonal terms of matrix $\tilde{\mathbf{C}}$ are computed as:

$$\tilde{C}_{pp} = \frac{1}{\frac{\partial(\Delta h_{pp})}{\partial q_p}} \tag{2.20}$$

Considering the Hazen Williams formula, and $|q_{ij}| = G_{ij}^{0.54}|h_i - h_j|^{1-0.46}$ (Equation 2.6),

$$\tilde{C}_{pp} = \frac{1}{1.852R_{ij}|q_{ij}|^{0.852}} = \frac{1}{1.852}{G_{ij}}^{0.54}|h_i - h_j|^{-0.46} \tag{2.21}$$

It follows that the improved version of Equation 2.14 is:

$$\Delta\mathbf{h} = -\left(\mathbf{B}\tilde{\mathbf{C}}\mathbf{B}^T\right)^{-1}\Delta\mathbf{d} = -1.852\left(\mathbf{B}\mathbf{C}\mathbf{B}^T\right)^{-1}\Delta\mathbf{d} \tag{2.22}$$

Considering the Hazen Williams formula and no singular head loss.

2.3 Calibration

Shamir and Howard [1977] stated that calibration "consists of determining the physical and operational characteristics of an existing system and determining the data [that] when input to the computer model will yield realistic results". The American Water Works Association Research Comittee on Distribution Systems [1974] used the word "verified" in place of "calibrated" but described a process of calibration: "System simulation is considered verified during preliminary analysis for design when calculated pressures are satisfactorily close to observed field gage readings for given field source send-out and storage conditions. If simulation is not satisfactory, the possibility of local aberrations, such as open boundary valves, is investigated. In the absence of other expected causative factors, the assumed local arterial network loads are adjusted until computed and observed field pressures are within reasonable agreement for various levels and extremes of demand, pumping, and storage". Walski [1983] proposed a more precise definition: "Calibration of a water distribution model is a two step process consisting of: (1) Comparison of pressures and flows predicted with observed pressures and flows for known operating conditions (i.e., pump operation, tank levels, pressure reducing valve settings); and (2) adjustment of the input data for the model to improve agreement between observed and predicted valves. A model is considered calibrated for a set of operating conditions and water uses if it can predict flows and pressures with reasonable agreement".

Savic et al. [2009] concluded from the literature review that a high degree of interest in this topic has been shown by researchers, but it has been considerably less covered by practitioners. A number of questions have to be answered, such as: (1) What parameters can be calibrated with confidence? (2) What is the acceptable level of discretization of calibration parameters and what is the acceptable level of agreement between measurements and model outputs? (3) How to parametrize the model when insufficient data are available? and (4) What objective function type to use?

Ormsbee [1989] suggested a seven-step general calibration procedure as follows: (1) Identification of the intended use of the model; (2) determination of initial estimates of the model parameters; (3) collection of calibration data; (4) evaluation of the model results; (5) macro-level calibration; (6) sensitivity analysis; and (7) micro-level calibration.

One of the most important issues in model calibration is the determination of the purpose of the model (Walski [1995]). Seven possible purposes of a network

model were identified: pipe sizing for master planning, extended-period simulations for planning studies, subdivision layout, rehabilitation studies, energy usage studies, water quality models and flushing programmes. Sumer and Lansey [2009a] modeled a real system for daily pump scheduling and system expansion design to examine the impact of model purpose on the calibration process.

Ostfeld et al. [2012] provides a summary of the Battle of the Water Calibration Networks (BWCN), the goal of which was to objectively compare the solutions of different approaches to the calibration of water distribution systems through application to a real water distribution system. Interesting references have been extracted from this work and future work is well pointed:

- Due to the inherently "ill posed or underconstrained" calibration problem in WDN, the solutions that provide a good match between measured and modeled data have to be validated with extra data.
- Uncertainty has to be included in the model parameters to explore the influence on the calibrated model outputs.
- Calibration size problem reduction is an important factor to consider to avoid model over-fitting, avoid unnecessary simulations or reducing the search space.
- Leakage data may be included in hydraulic calibration efforts because leakage directly affects nodal demand allocation and pump curve characterizations.
- The effect of different field data on model calibration should be investigated (use of flow and/or pressure measurements).

In Walski [1985], the author described the importance of good data collection. In Walski [2000], the same author classified data into three different degrees of usefulness:

- Good data are collected when there is sufficient head loss to draw valid conclusions about model calibration. Walski et al. [2014] stated that it is necessary to have head loss in the system that is significantly greater than the error in measurement to avoid random adjustments.
- Bad data contain errors because of misread pressure gauge, incorrectly determined elevation of the pressure gauge, or lack of information about which pumps were running when calibration data were collected. This type of data should be discarded.

- Useless data are collected when the head loss in the system is so low that head loss and velocity are of a similar order of magnitude as the errors in measurements. Such data can produce misleading models.

Ahmed et al. [1999] developed a heuristic three-step procedure to assist in identifying the conditions under which useful data (good data) should be collected. The issue of data quality and quantity is closely related to that of sampling design, which will be addressed later in this chapter.

Walski [1983] presented formulas to assist the user in deciding whether to adjust roughness or water use and by how much. They are based on fire flow test. To correct for inaccuracies in input data it is necessary to first understand the sources of these inaccuracies. These can be grouped into several categories: (1) Incorrect estimate of water use; (2) incorrect pipe carrying capacity; (3) incorrect head at constant head points (i.e., pumps, tanks, pressure reducing valves); or (4) poor representation of system in model (e.g., too many pipes removed in skeletonizing the system). The major source of error in simulation of contemporary performance will be in the assumed loadings distributions and their variations. On the other hand, Eggener and Polwoski [1976] state: "the weakest piece of input information is not the assumed loadings condition, but the pipe friction factor". The certainties of a previous model must be stated so that the effort in calibration is in the good direction.

Goulet et al. [2013] assessed that the most important uncertainty sources are demands and model simplifications, but uncertainty also originates from measurement errors, incorrect boundary conditions, inherent model structural errors or unknown status of valves [Hutton et al. [2014], Walski et al. [2014]]. The calibration in this work focuses on demands due to their daily variability and continuous evolution depending generally on social and climate factors comparing to the more stable evolution of roughness.

In projects developed by the doctoral candidate's research group (PROFURED, RTNM, EFFINET), the experts assessed that one of the main causes of uncertainty in the models, and consequently in results, were the nodal demands Pérez et al. [2011a]. The present work focuses in calibrating these demands, which cause the major prediction errors in the studied cases in the projects named before.

2.3.1 The unknown inputs: Demands

Demands are not physically in the network like nodes or pipes. They are inputs because they are the driving force behind the hydraulic dynamics occurring in

water distribution systems Walski et al. [2003]. Any place where water can leave the system represents a point of consumption, including a customer's faucet, a leaky main, or an open fire hydrant.

Three questions related to water consumption must be answered when building a hydraulic model: (1) How much water is being used? (2) Where are the points of consumption located? and (3) How does the usage change as a function of time?

Determining demands is not a straightforward process like collecting data on the physical characteristics of a system. Some data, such as billing and production records, can be collected directly from the utility but are usually not in a form that can be directly entered into the model. Once this information has been collected, establishing consumption rates is a process requiring study of past and present usage trends and, in some cases, the projection of future ones. The bottom-up approach for determining the demand model requires the study of the socioeconomic reality of the consumer (Arbués et al. [2003]), or the study of particular uses of water (Creaco et al. [2015]). Piller and Bremond [2002] presented a stochastic demand model for peak period analysis useful for network design.

However, the top-down approach is the most commonly used. The model presented in Equation 2.12 represents this approach, where the demand multiplier pattern observed at the network inputs (total consumption) is applied to all network nodes, considering a correction factor that depends on the average demand value of the generic node. Ideally, if individual meter readings are taken for every customer, they should exactly equal the amount of water that is measured leaving the treatment facility. In practice, however, this is not the case. Although inflow does indeed equal outflow, not all of the outflows are metered. These lost flows are referred to as unaccounted-for-water (UFW). Leakage is frequently the largest component of UFW and includes distribution losses from supply and distribution pipes, trunk mains, services up to the meter, and tanks. The amount of leakage varies depending on the system, but there is a general correlation between the age of a system and the amount of UFW. Newer systems may have as little as 5% leakage, while older systems may have 40% leakage or higher. Leakage tends to increase over time unless a leak detection and repair program is in place. There are some methodologies to study the UFW by means of the minimal night flow (Lambert [1994]) and the DMA performance. If better information is not available, UFW is usually spread uniformly around the system (in spatial and time terms). If UFW is reduced, then the utility will see higher peaking factors

because UFW tends to flatten out the diurnal demand curve.

Although water utilities make a large number of flow measurements, such as those at customer meters for billing and at treatment plants and wells for production monitoring, data are usually not compiled on the node-by-node basis needed for modelling. The modeller is thus faced with the task of spatially aggregating data in a useful way and assigning the appropriate usage to model nodes. The most common method of allocating base demands is a simple unit loading method. Most modellers start by determining base demands to which a variety of peaking factors and demand multipliers can be applied, or to which new land developments and customers can be added. Base demands typically include both customer demands and unaccounted-for-water. Usually, the average day demand in the current year is the base from which other demand distributions are built (Equation 2.12 and Equation 2.11). Ideally, the process of loading demand data into a model from another source would be relatively automatic. Cesario and Lee [1980] described an early approach to automate model loading.

Water usage in municipal water distribution systems is inherently unsteady due to continuously varying demands. For an extended period simulation to accurately reflect the dynamics of the real system, these demand fluctuations must be incorporated into the model. The temporal variations in water usage for municipal water systems typically follow a 24-hour cycle called a diurnal demand pattern. However, system flows change not only on a daily basis, but also weekly and annually. As one might expect, weekend usage patterns often differ from weekday patterns. Seasonal differences in water usage have been related to climatic variables such as temperature and precipitation, and also to the changing habits of customers, such as outdoor recreational and agricultural activities occurring in the summer months.

Demands in a scenario are not the result of real-time measurements, but instead consist of educated guesses that can be derived from a number of sources such as typical usage by consumers, customer billing records, and required fire fighting loads. To be effective a real-time modeling technique must acknowledge and accommodate the disconnection between mean demand estimates that change gradually and real demands (Davidson and Bouchart [2006]).

The adjustment of demands is a typical inverse problem based on optimization. Aksela and Aksela [2011] used a weekly consumption calculated and measured for classification of different households. This classification allows an estimation of the curves using Gibbs sampling and combined Gaussians. Davidson and Bouchart [2006] and Cheng and He [2011] described how to estimate de-

mands using head and flow measurements by means of explicit methods based on least squares (LS). Both formulate the problem for networks where the number of measurements and the number of parameters to estimate are similar.

2.3.2 Calibration methods

Global calibration problem is very well presented by Savic et al. [2009]. Methods are classified depending on their dynamics (static/transient) and depending on the optimization methods (prove/explicit/implicit).

Iterative calibration models (Rahal et al. [1980], Walski [1983, 1986], Bhave [1988]) are based on trial-and-error procedure. Unknown parameters are updated at each iteration using heads and/or flows obtained by solving the set of steady-state mass balance and energy equations. Iterative calibration models have been the base in the establishment of some fundamental principles and guidelines regarding WDS model calibration, and have been utilized in the development of more sophisticated explicit and implicit methods.

Explicit calibration models (Ormsbee and Wood [1986], Boulos and Wood [1990], Ferreri et al. [1994]) are based on solving an extended set of steady-state mass-balance and energy equations. This extended set is solved explicitly, usually by the Newton Raphson method. The main disadvantages and limitations are: (1) The number of parameters to be calibrated must be equal to the number of measurements, (2) measurements errors are not taken into account, and (3) there is no way to quantify uncertainty in the estimated parameters. In conclusion, explicit methods only have historical significance and no apparent influence on the current practice of model calibration (Savic et al. [2009]).

Implicit methods are formulated and solved using an optimization technique coupled with a hydraulic solver. The optimization tool sets/updates parameters, passes them onto the simulation model, which in turn passes back obtained model predicted variables. The optimization tool employs an objective function to minimize the differences between measured and model predicted variables. The types of optimization methods used vary from local search methods, through mathematical optimization to global search methods.

Non-evolutionary optimization methods

Multiple types of non-evolutionary methods exist, but gradient-type optimization seems to be dominant. General Reduced Gradient (GRG) is used in Lansey and Basnet [1991] and Shamir [1974]. The Gauss-Newton method, and the improved

version of Levenberg-Marquardt, are the most used gradient type methods. Datta and Sridharan [1994], Pudar and Liggett [1992], Liggett and Chen [1994], Reddy et al. [1996], Piller et al. [1999], Lansey et al. [2001], Davidson and Bouchart [2006], Kang and Lansey [2009], Cheng and He [2011] and Kang and Lansey [2011] used these methods to solve the linear/non-linear least squares formulation of the inverse problem. The influence of different choice of weights in the WLS case is investigated and a systematic procedure is given for the selection of suitable weights in Datta and Sridharan [1994] and Reddy et al. [1996]. The non-linear least squares problem is a non-convex problem with multiple optima for the objective function, and it is generally solved as an iterative procedure. To be certain that the minimum found is the global minimum, the process should be started with widely differing initial values of the parameters. When the same minimum is found regardless of the starting point, it is likely to be the global minimum. Other non-evoluationary techniques used for calibration of hydraulic distribution models include the extended complex method of Box (Ormsbee [1989]), linear and non-linear programming (Greco and Del Giudice [1999]), Kalman filtering (Todini [1999]), and simulated annealing (Tucciarelli et al. [1999]).

Evolutionary optimization methods

Evolutionary optimization methods (Goldberg [1989]) were introduced in the area of calibration of WDN models by Savic and Walters [1995]. Genetic Algorithms (GA) in particular, have been used later on in hydraulic models calibration works (Walters et al. [1998], Lingireddy and Ormsbee [1999], Tang et al. [1999], Vítkovský et al. [2000], Kapelan [2002]).

Savic et al. [2009] pointed out that the main advantages of the evolutionary methods over non-evolutionary methods are the lack of complex mathematical apparatus to evaluate sensitivities or invert matrices, the ability to handle with large calibration problems, easy incorporation of additional calibration parameter types and constraints, and the opportunity to implement in multiple computational machines due to their parallel nature. However, Walters et al. [1998] noted some disadvantages: not guaranteeing the achievement of the global optimum in large and complex systems, need for a careful setting up and parameter tuning to obtain a correct operation, and reduced computational efficiency when compared to the gradient-based methods. Consequently, they can be used for large optimization problems but not in real time. Furthermore, evolutionary methods are more robust in finding the global optimum in non-convex optimization problems.

Maier et al. [2014] clarifies the current status and future research directions of evolutionary algorithms and other meta-heuristics for better solving key water resources problems.

2.4 Identifiability

The inverse problem is often ill-posed. The ill-posedness is generally characterized by the non-uniqueness of the identified parameters. The uniqueness problem in parameter estimation is intimately related to identifiability (Yeh [1986]).

Observability and identifiability terms are sometimes confound. System observability determines if the state of a system, i.e. the system variables (head, flow), can be estimated. On the other hand, system identifiability resolves if the parameters of the system (consumptions, roughness's coefficients) can be calibrated. In conclusion, observability refers to system state (dynamic variables) while identifiability refers to system parameters (assumed constant in a certain time horizon).

An important contribution to the solution of the observability problem was made by Krumpholz et al. [1980], who formulated necessary and sufficient conditions for observability in power-system state estimation in terms of meter location and network topology. According to their analysis a network is observable if and only if it contains a spanning tree of full rank. Bargiela [1985] formulated the same problem for water systems.

Pérez [2003] classified the identifiability as static and dynamic. Carpentier and Cohen [1991] performed the study of identifiability for the static problem using graph analysis based on Ozawa [1987]. The idea is that some operations in graphs are equivalent to operation on equations.

Conditions of identifiability for non-linear dynamic systems can be found in the literature. Walter and Pronzato [1996] used the state space formulation by means of the dynamic information of the system. For the linear case, the invertibility of the matrix of the equations set was studied by Sorenson [1980].

The complexity of the transient equations in dynamic identifiability makes their use difficult for real networks. The extended period identifiability is based on quasi-static equations, which allows use of simpler equations related from one time step to the next one by tank equations. Pérez [2003] studied the extended period identifiability based on the sensitivity matrix rank in both linear and non-linear cases. The author stated that if many measurements are taken in the

same conditions they will not add any information (not increase the rank of the sensitivity matrix) but could be useful for filtering the noise in the measurements.

2.5 Uncertainty

In calibration, inaccuracy of the input data causes the results to be inaccurate too. Therefore, it is important to give not only the estimated values of the calibration, but also an indication of how reliable these estimations are.

Hutton et al. [2014] presented a framework where each stage of model development is considered, and reviewed the most promising methods available to quantify and reduce uncertainty at each of these stages (calibration, data assimilation and model forecasting). Sumer and Lansey [2009b] discussed linking the impact of calibration uncertainties to model decisions and examining the implications of the uncertainties on future data collection efforts.

Bargiela and Hainsworth [1989] presented and compared three methods for confidence limit analysis:

- Monte Carlo simulation: Uncertainty in model predictions is calculated by a series of simulations where the input parameter's vector has random variations.
- Optimization-based approach: The confidence limits of the estimated values are calculated by means of an optimization problem with the linearized network equations as constraints.
- Sensitivity-based method: Analysis of the sensitivity matrix generated from the linearized network equations.

The authors selected the latter approach as the better one due to the improvement on the computational requirements keeping similar results as the other methods. Furthermore, most of the reviewed bibliography (Yeh [1986], Datta and Sridharan [1994], Reddy et al. [1996], Bush and Uber [1998], Ahmed et al. [1999], Piller et al. [1999], Lansey et al. [2001], Mallick et al. [2002], Kapelan et al. [2003b, 2005], Behzadian et al. [2009], Kang and Lansey [2009]) perform the quantification of the parameter and prediction uncertainties based on linear regression theory, a method known in the literature as the FOSM model (First-Order Second-Moment) (Bard [1974]).

The FOSM model consists in the definition of the first-order approximation of the parameter covariance matrix $\mathbf{Cov}(\hat{\mathbf{x}})$ as

$$\mathbf{Cov}(\hat{\mathbf{x}}) = \sigma^2(\mathbf{S}^T\mathbf{S})^{-1} \tag{2.23}$$

where σ^2 is the variance in measured observations, and $\mathbf{S}$ is the matrix of the sensitivities of the measures relative to the estimated parameters $\hat{\mathbf{x}}$. Uncertainty in the parameter values is indicated by parameter variances in the ith diagonal element of the covariance matrix.

The prediction covariance matrix $\mathbf{Cov}(\hat{\mathbf{y}})$ can be also estimated to obtain the variance of the model prediction:

$$\mathbf{Cov}(\hat{\mathbf{y}}) = \mathbf{S_P} \cdot \mathbf{Cov}(\hat{\mathbf{x}}) \cdot \mathbf{S_p}^T \tag{2.24}$$

where $\mathbf{S_P}$ is the matrix of the sensitivities of the predicted values relative to the estimated parameters $\hat{\mathbf{x}}$.

We can compute 95% confidence intervals for the individual model parameters considering that each model parameter $\mathbf{x}_i$ has a normal distribution with mean $\hat{\mathbf{x}}_i$ and variance $\mathbf{Cov}(\hat{\mathbf{x}})_{i,i}$. The 95% confidence intervals are given by

$$\hat{\mathbf{x}}_i \pm 1.96 \cdot \sqrt{\mathbf{Cov}(\hat{\mathbf{x}})_{i,i}} \tag{2.25}$$

where the 1.96 factor is obtained from

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-1.96\sigma}^{1.96\sigma} e^{-\frac{x^2}{2\sigma^2}} dx \approx 0.95 \tag{2.26}$$

This information allows us to represent the probability density function (PDF) of the calibrated parameters.

2.5.1 Optimality

The uncertainty propagated to the calibrated parameters depends on the information used to calibrate them, i.e. the sampling design. The A, D and V-optimality criteria are used to evaluate how well a sampling design represents the true behavior ('response') of a WDS. The A-optimality minimizes the average parameter variance; the D-optimality maximizes the determinant of the curvature matrix (Kapelan et al. [2003b]); and the V-optimality minimizes the average prediction variance (Savic et al. [2009]). The optimality of a set of sensors can be analyzed by

means of the formulas proposed by Kapelan et al. [2003b], listed in Equation 2.27.

$$F_1 = \frac{1}{n_p}\sum_{i=1}^{n_p}\mathbf{Cov}(\hat{\mathbf{x}})_{i,i}^{1/2} \qquad F_2 = [det(\mathbf{S}^T\mathbf{WS})]^{1/(2n_p)} \qquad F_3 = \frac{1}{n_z}\sum_{i=1}^{n_z}\mathbf{Cov}(\hat{\mathbf{y}})_{i,i}^{1/2} \tag{2.27}$$

where n_p is the number of parameters, and n_z is the number of predicted variables of interest, i.e. the number of chosen predictions whose uncertainties are being evaluated. F_1 and F_2 are based on the parameter uncertainty (A-optimality and D-optimality, respectively), while F_3 is based on the model prediction uncertainty (V-optimality).

2.6 Sampling design

Calibration accuracy should be judged both by the model's ability to reproduce data, and by a quantitative measure of the uncertainty in calibrated parameter values. This uncertainty depends on the sampling design, including the measurement type, number, location, frequency, and conditions existing at the time of sampling (Bush and Uber [1998]).

Reviewed literature defines the sampling design as the procedure to determine (Kapelan et al. [2003b]): (1) what WDS model predicted variables (pressures, flows, both, etc.) to observe; (2) where in the WDS to observe them; (3) when to observe (in terms of duration and frequency); and (4) under what conditions to observe.

In general, a sampling design may have one of several purposes (Loaiciga et al. [1992]): Ambient monitoring, detection, compliance, or research. Model calibration is considered research sampling, where the objective is to identify accurately the physical parameters of the system. A sampling design, $\mathbf{s}$, is a set of specified measurements $\mathbf{y}$, at particular locations and times, along with the experimental conditions under which measurements are made (Bush and Uber [1998]).

Walski [1983] proposed one of the first sampling designs by suggesting to: (1) Monitor pressure near the high demand locations; (2) conduct fire flow tests on the perimeter of the skeletal distribution system, away from water sources; (3) use as large as possible test flow at the fire hydrant; and (4) collect both head and flow measurements.

Liggett and Chen [1994] impressed the importance of sensitivity in inverse problems for two primary reasons. First, the need for the measurements to be

made at a location where they are sensitive to the desired calibration parameters. Second, the degree of confidence that one has in the result depends on the sensitivity.

Different approaches for solving the optimization problem have been developed. Usually, the main objective of finding the best locations for sensors is combined with other objectives (i.e. devices' cost). GAs, sensitivity matrix analysis or heuristic methods are some of the methodologies used.

Yu and Powell [1994] formulated the meter placement problem as a multiobjective optimization by seeking the best solution in terms of estimation accuracy and metering cost. They developed a method employing dynamic analysis of the covariance matrix of state variables and the decision-trees technique.

Ferreri et al. [1994] ranked the potential location of the sensors according to their overall relative sensitivity of nodal heads with respect to roughness coefficients. Bush and Uber [1998] proposed three general sensitivity-based methods derived from the D-optimality criterion to rank the locations and types of measurements for estimating the roughness coefficients of a WDS model using pressure measurements, tracer concentration measurements and a combination of both. The authors outlined that the proposed methods, although suboptimal, may have some advantages over purely statistical methods that lack a physical basis. Del Giudice and Di Cristo [2003] compared these three sensitivity-based methods for selecting the worthwhile pressure and flow sensors location in water distribution network for calibrating roughness coefficients. Ahmed et al. [1999] developed a sensitivity-based heuristic method (also derived from the D-optimality criterion), to study the uncertainty caused by measurement and estimation errors in WDN.

Pinzinger et al. [2011] proposed three algorithms based on integer linear programming and Greedy paradigm. Piller et al. [1999] formulated the sampling design (SD) as an optimization problem which minimizes the influence of measurement errors in the state vector estimation subject to the constraint that the Jacobian matrix is of maximum rank. A greedy algorithm was used, which selected at each iteration the optimal location of the sensors.

Some of the mentioned approaches used an iterative selection of the sensors, adding at each iteration one sensor to the set of already located ones. However, Kapelan et al. [2003b] demonstrated that the optimal set of locations for n monitoring points is not always a superset of the optimal set for $n-1$ monitoring points.

Lansey et al. [2001] developed a sensitivity-based heuristic sampling design

procedure for WDS model calibration to identify preferable conditions for data collection, accounting for uncertainty in measurements and its impact on both model parameters and predictions.

de Schaetzen et al. [2000] proposed three sampling design approaches. The first two were based on the shortest path algorithm, and set sensors' locations depending on the distance between the source and the set of potential sensors nodes. The third approach solved the optimization problem based on maximization of Shannon's entropy, locating sensors in the nodes with highest pressure sensitivity on roughness changes. The sampling design cost was also taken into account.

Meier and Barkdoll [2000] introduced GA in sampling design to find the combination of fire-flow test locations that, when analyzed collectively, stresses the greatest percentage of the hydraulic network, so the roughness parameters of grouped pipes can be calibrated. Kapelan et al. [2003b] and Kapelan et al. [2005] presented a multi-objective sensitivity-based method for sampling design where both uncertainty and SD cost objectives where minimized. Model accuracy was maximized and formulated as the D-optimal criterion, the A-optimal criterion and the V-optimal criterion. SOGA/MOGA (Single/Multi Objective GA) were used and compared, leading to the conclusion that the advantages in MOGA outweigh its disadvantages. The Jacobian matrix used was calculated prior to the optimization model run by assuming the model parameter values. Opposed to this deterministic approach, Behzadian et al. [2009] tried to overcome this latter assumption by introducing parameter uncertainty using some pre-defined probability density function. Results in studied cases (Kapelan et al. [2003a] and Kapelan et al. [2005]) assessed that the calibration accuracy based on prediction uncertainty (V-optimality) is preferred over parameter uncertainty (D-optimal and A-optimal criteria). Similarly, D-optimality is preferred over A-optimality.

Kang and Lansey [2010] posed the sampling design as a multiobjective optimization problem where the objective functions represented demand estimation uncertainty, pressure prediction uncertainty and demand estimation accuracy. The optimization problem was solved using MOGA based on Pareto-optimal solutions.

Not all sampling design approaches are addressed to parameter calibration. Pérez et al. [2009] proposed a sampling design based on a leakage detection methodology. One sensor was located at each iteration of the procedure with the objective of minimizing the maximum number of nodes with the same binary signature (which cannot be isolated separately). Nejjari et al. [2015] developed an

optimal sensor placement strategy based on pressure sensitivity matrix analysis and an exhaustive search strategy. Pérez and Sanz [2014] presents a comparison of different sensor placement methodologies for demand calibration and leak detection. Piller et al. [2015] investigates which performance criteria should be considered to place water quality and quantity sensors for both early detection and model calibration.

2.7 Fault detection

Waste and loss of water have been sometimes disregarded due to the low water price and ease of exploitation in developed countries. However, both users and utilities are increasing their concern to avoid present and future water scarcity. Individual users can optimize their daily routines to reduce water waste, but burst and background leakage will be present independently of it.

Leakage in water distribution systems has attracted a lot of attention by both practitioners and researchers over the past years. Puust et al. [2010] provides a review of leakage management related methods in distribution pipe systems from detection and assessment to efficient control. Leakage identification is divided into leakage awareness and leakage localization (Puust et al. [2010]). Leakage awareness focuses on leakage detection in the network (Kapelan et al. [2003c], Mounce et al. [2010, 2011], Palau et al. [2012], Romano et al. [2014]), but does not give any information about its precise location. On the other hand, leakage localization (Romano et al. [2013]) is an activity that identifies and prioritizes the areas of leakage to make pinpointing of leaks easier. Leak localization techniques can be divided into two categories: external and internal (ADEC [1999]). The use of external methods like acoustic logging (Pilcher [2007]), penetrating radar (Hugenschmidt and Kalogeropoulos [2009]) or liquid detection methods (Henault et al. [2010]) has some drawbacks like needing a large number of sensors, not being suitable for application in large urban areas, or being invasive. Internal methods use continuously monitored data to infer the position of leaks using models. Many techniques can be found in literature (Liggett and Chen [1994], Vítkovský et al. [2000], Kim [2005], Colombo et al. [2009]). All of these techniques are based on transient analysis, which is mainly used on single, grounded pipelines due to the high effect of the system uncertainty on results. Non-transient model-based leakage localization techniques have been also developed during the last years (Wu and Sage [2006], Pérez et al. [2011b], Wu et al. [2010], Farley et al.

[2011], Goulet et al. [2013], Pérez et al. [2014]). These techniques analyze the difference between measurements and estimated values from leaky scenarios to signal the probability of a zone to contain leakage. Some of these model-based methodologies assume the hypothesis of a single leak in the network (Goulet et al. [2013], Pérez et al. [2014]). Wu et al. [2010] calibrated leakage as a pressure driven demand using the competent genetic algorithm, providing a tool for assisting leakage detection engineers to predict leakage hot-spots. Walski et al. [2014] provide some practical suggestions to help users collect the right quality and quantity of data and interpret the results when running GA to locate leaks and incorrectly closed valves. Wu and Song [2012] have developed an efficient method to effectively locate the known valves and identify not only their status but also the settings.

Chapter 3

Sensitivity analysis for parameterization and sampling design

This chapter describes the core of the developed thesis: the demand components model. First, a review of the background theory of the singular value decomposition (SVD) is performed. Next, the whole concept of the demand components model, and the justification of its use, are presented. Then, the analysis of the SVD of the sensitivity matrix is used to define the demand components that will be calibrated. Finally, the same analysis is performed to select the type and location of the sensors that will be used to take measurements from the water network. A dummy network is used at each step to exemplify each of the processes.

3.1 Background theory review: the singular value decomposition

The SVD is a matrix decomposition method whereby a general m by n system matrix $\mathbf{A}$, relating parameters $\mathbf{x}$ to observations $\mathbf{y}$:

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{y}, \tag{3.1}$$

is factored into

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{V}^T, \tag{3.2}$$

Table 3.1: Example of a sensitivity matrix

-0,039	-0,012	-0,005	-0,039	-0,005	-0,005	-0,005	-0,039	-0,039	-0,005
-0,012	-0,107	-0,049	-0,012	-0,049	-0,049	-0,049	-0,012	-0,012	-0,049
-0,005	-0,049	-0,059	-0,005	-0,059	-0,059	-0,059	-0,005	-0,005	-0,059
-0,039	-0,012	-0,005	-0,778	-0,005	-0,005	-0,005	-0,778	-0,778	-0,005
-0,005	-0,049	-0,059	-0,005	-4,857	-4,100	-1,950	-0,005	-0,005	-4,100
-0,005	-0,049	-0,059	-0,005	-4,100	-4,100	-1,950	-0,005	-0,005	-4,100
-0,005	-0,049	-0,059	-0,005	-1,950	-1,950	-1,950	-0,005	-0,005	-1,950
-0,039	-0,012	-0,005	-0,778	-0,005	-0,005	-0,005	-1,362	-1,362	-0,005
-0,039	-0,012	-0,005	-0,778	-0,005	-0,005	-0,005	-1,362	-1,605	-0,005
-0,005	-0,049	-0,059	-0,005	-4,100	-4,100	-1,950	-0,005	-0,005	-5,504

where $\mathbf{U}$ is a set of m orthonormal singular vectors that form a basis of the observed data vectorial space, $\mathbf{V}$ is a set of n orthonormal vectors that form a basis of the parameter vectorial space, and $\mathbf{\Lambda}$ is an m x n diagonal matrix of singular values of $\mathbf{A}$, where the additional rows (more measurements than parameters) or columns (more parameters than measurements) are filled with zeros (Aster et al. [2005]).

The SVD has many applications that can be useful for the main purpose of this thesis: demand calibration. The key step to ensure the success of the calibration is the grouping of nodal demands into fewer parameters that, in the end, keep the network behavior as close to the original behavior as possible. This grouping ensures the identifiability of the system.

When calibrating parameters in non-linear systems (Chapter 4), the system matrix $\mathbf{A}$ in Equation 3.1 is replaced by the system sensitivity matrix $\mathbf{S}$, which relates changes in data with changes in parameters. Explanations from now on focus on the sensitivity matrix $\mathbf{S}$.

Consider the sensitivity matrix $\mathbf{S}$ in Table 3.1, obtained from the network in Figure 3.7. After performing the SVD of matrix $\mathbf{S}$, we can compute a reconstructed matrix $\mathbf{S_r}$ from a subset of columns of $\mathbf{U}$ and $\mathbf{V}$, ignoring the information from these matrices that correspond to low relevant singular values. The singular values in matrix $\mathbf{\Lambda}$ are depicted in Figure 3.1. We can observe very low singular values from the fifth position, indicating that the corresponding columns in matrices $\mathbf{U}$ and $\mathbf{V}$ have low importance in the reconstruction of matrix $\mathbf{S}$. The consideration of how low is a particular singular value is extracted from which contribution it does to the reconstruction of matrix S (see Figure 3.1 and Figure 3.2), which is related to the ratio between the highest and the i^{th} singular value. Figure 3.2 presents the mean squared error (MSE) in the reconstruction of the sensitivity matrix depending on the number of columns used from matrices

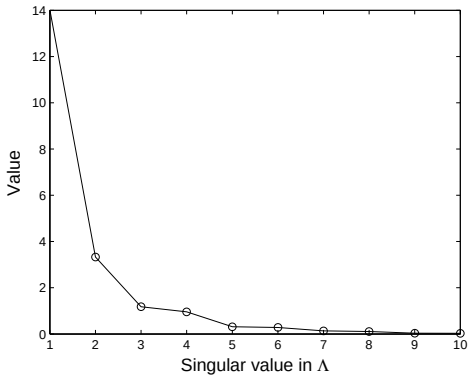

Figure 3.1: λ singular values from the SVD of the sensitivity matrix in Table 3.1

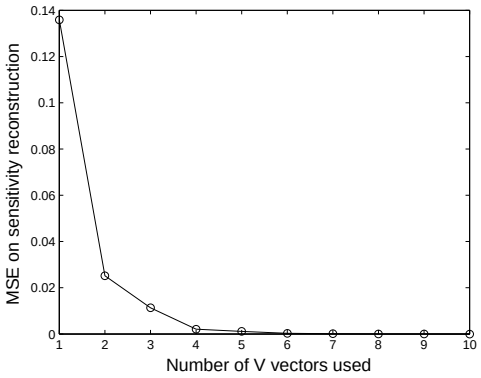

Figure 3.2: MSE of the reconstructed matrix S from a different number of **U** and **V** vectors

Table 3.2: Reconstructed sensitivity matrix

-0,001	-0,001	0,000	-0,026	-0,005	-0,006	-0,006	-0,041	-0,045	-0,005
-0,001	-0,002	-0,002	-0,008	-0,048	-0,050	-0,053	-0,013	-0,014	-0,049
0,000	-0,002	-0,002	-0,004	-0,058	-0,061	-0,063	-0,006	-0,007	-0,059
-0,026	-0,008	-0,004	-0,522	-0,005	-0,006	-0,006	-0,822	-0,887	-0,005
-0,005	-0,048	-0,058	-0,005	-4,788	-4,216	-1,929	-0,005	-0,006	-4,071
-0,006	-0,050	-0,061	-0,006	-4,216	-3,906	-1,985	-0,005	-0,005	-4,149
-0,006	-0,053	-0,063	-0,006	-1,929	-1,985	-1,944	-0,005	-0,005	-1,941
-0,041	-0,013	-0,006	-0,822	-0,005	-0,005	-0,005	-1,295	-1,397	-0,005
-0,045	-0,014	-0,007	-0,887	-0,006	-0,005	-0,005	-1,397	-1,508	-0,005
-0,005	-0,049	-0,059	-0,005	-4,071	-4,149	-1,941	-0,005	-0,005	-5,492

U and **V**, corresponding to the same number of singular values in $\mathbf{\Lambda}$, computed as:

$$MSE_k = \frac{\sum_i^{n_{row}} \sum_j^{n_{col}} (\mathbf{S_o}_{(i,j)} - \mathbf{S_r}_{(i,j)})^2}{n_{row} + n_{col}} \qquad k = 1..10 \tag{3.3}$$

where n_{row} and n_{col} are the number of rows and columns in both the original ($\mathbf{S_o}$) and reconstructed ($\mathbf{S_r}$) sensitivity matrices, and equal to 10 in this example. It can be seen that when considering only the four first columns the MSE falls to a very low value. Table 3.2 presents the reconstructed sensitivity matrix in this particular case.

The reduction of matrices **U** and **V** is used in the methodology presented to choose which parameters will be calibrated, and which sensors will be used in the calibration process. In this thesis, the number of singular values considered, and consequently the number of **U** and **V** columns used, depends on the number of parameters and sensors that will be defined. Wiggins [1972] suggests a threshold value of 10^{-2}, and Wasantha Lal [1995a] a threshold value of 10^{-3}. The results presented do not reach these low values, as the number of sensors used is limited.

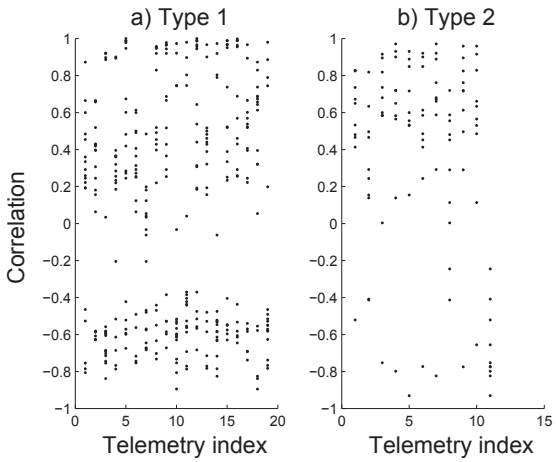

Figure 3.3: Cross correlations of Nova Icària DMA telemetries in users of segments a) 1; and b) 2

3.2 Demand components model

In Chapter 2 we have seen that the most used demand models are the basic demand model (Equation. 2.12) and the demand patterns model (Equation 2.11). The demand model presented in Equation 2.12 cannot explain the daily variation in the relative pressure behavior between two areas in the network, as it fixes the same behavior to all demands. On the other hand, the demand patterns model requires a lot of information that is not usually available (users associated with a given node, type of users), or does not fulfill the assumptions (incorrect predetermined diurnal demand patterns' values, users of the same type behaving differently). An example of the latter is presented in Figure 3.3: telemetries from two different segments (i.e. types of users) from a real network called Nova Icària in Barcelona, have been analyzed. Each telemetry consists of the daily water consumption of a specific user, metered hourly. The correlation between every pair of telemetries within the same segment has been computed to assess the distance between their profiles, i.e. the similarity or dissimilarity of the users' behaviors. In each subfigure, the x-axis presents the users' telemetries, and each dot in the y-axis indicates the correlation between the user and all the other users in the same segment: the higher the correlation, the higher the similarity between two telemetries. Figure 3.3.a presents a type of user with no similarity between its members, whereas Figure 3.3.b shows a type of user with more similarity between its members, but not enough to assume that all of them behave in the same way. In conclusion, the assumption of considering that all users of the same

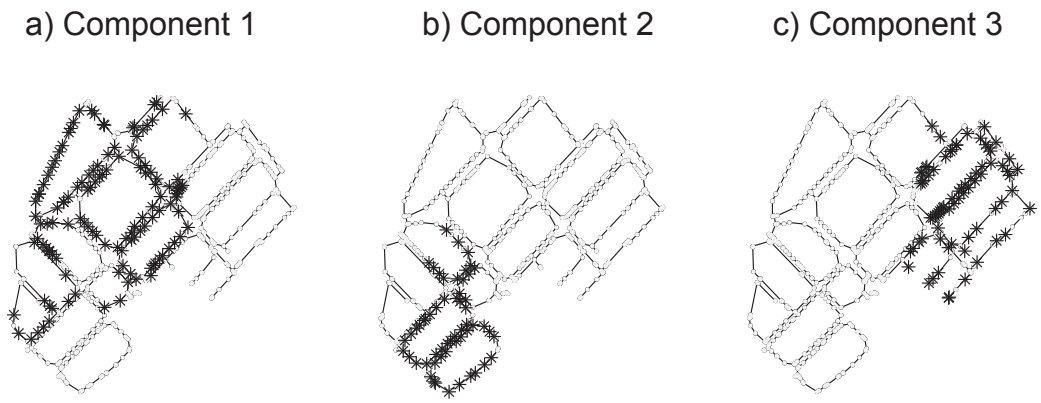

Figure 3.4: Example of demand components with binary memberships

type behave in the same way can lead to incorrect results or high uncertainty in the calibrated parameters.

This thesis proposes a new approach to model demands depending on their geographical location, and their sensitivity to hydraulic variables. Initially, nodes in a specific zone of the network were assigned to a specific behavior, which from now on, will be called demand component. Equation 3.4 presents this new model:

$$\mathbf{d}_i(t) = \frac{bd_i}{\sum_{j=1}^{n_d} bd_j} \cdot \mathbf{c}_{j \to i}(t) \cdot \mathbf{q_{in}}(t) \tag{3.4}$$

where $\mathbf{c}_{j \to i}(t)$ is the value of the demand component j associated to node i depending on the node location. Demand components are calibrated demand multipliers that represent the behavior of nodes in a determined geographical zone, avoiding the dependency on information of the user type and diurnal pattern behavior. All nodes in the same area of node i have the same associated demand component. Consequently, all nodes in the same zone will have the same demand behavior, weighted depending on their base demand. This demand model is capable of generating pressure variations in different zones of the network, as it happens in a real situation. Figure 3.4 presents a network where three demand components have been defined. Each subplot presents the set of nodes that are modulated by the same demand component (Equation 3.4).

However, the assumption that all nodes in the same area behave exactly in the same way is not realistic. For example, a node in the limit of the effect zone of two demand components should probably have a combination of the behavior of the two demand components, instead of only one. To solve that, we can

redefine the demand model in Equation 3.4 so that the degree to which each demand component is associated with each node is given as a membership, which depends on the nodes' geographical location. Equation 3.5 represents the new demand model:

$$\mathbf{d}_i(t) = \frac{bd_i}{\sum_{j=1}^{n_d} bd_j} \cdot \mathbf{q_{in}}(t) \cdot (\alpha_{i,1} \cdot \mathbf{c}_1(t) + \alpha_{i,2} \cdot \mathbf{c}_2(t) + \cdots + \alpha_{i,n_c} \cdot \mathbf{c}_{n_c}(t)) \quad (3.5)$$

with

$$\alpha_{i,1} + \alpha_{i,2} + \cdots + \alpha_{i,n_c} = 1 \quad \forall i$$

where $\alpha_{i,j}$ is the association of demand component j with node i, and n_c is the number of demand components. The membership $\alpha_{i,j}$ of each node to each demand component depends on the geographical location of the node, and is computed by means of the sensitivity analysis presented later in Section 3.3. The model in Equation 3.5 can generate different behaviors in every demand, while only having to calibrate few (n_c) demand components.

This way of calibrating demands incorporates the usually ignored fact that demands depend in some ways on the head status of the network (Giustolisi and Walski [2012]). For example, if the pressure in a specific zone of the DMA decreases, the calibration process will estimate demand component values that decrease the consumption of nodes in that zone. Demand components presented in this thesis should not be confused with the ones defined in Giustolisi and Walski [2012], where demand components were generated with a previous knowledge of the use of water (human-based, volume-based, non-controlled orifice-based, leakage-based).

The calibrated demand components generate individual demands that may not be exactly as the real ones, but the aggregated demand in a zone at a specific sample, and the cumulative demand of each individual node during a period of time (similar to the billing) should coincide with the real ones if other parameters (roughness, valve status, etc.) are well calibrated.

Figure 3.5 presents the nodal memberships to three demand components in an example network. The first component effect zone is located on the North-West side of the DMA (i.e. memberships of nodes to this component are higher in the North-West); the effect zone of the second component is located on the South-West of the DMA; and the third component effect zone is located on the East side of the network. The nodes' memberships are depicted in greyscale: the darker the color of a node, the higher the membership of that node to the demand

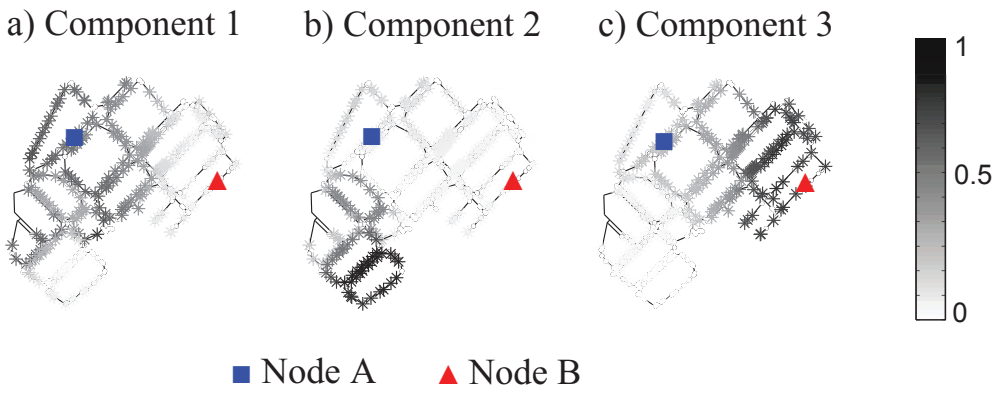

Figure 3.5: Example of demand components and memberships in a network

Table 3.3: Memberships of nodes A and B of the example network

Node	A	B
Membership to c_1	0.6	0.01
Membership to c_2	0.05	0.01
Membership to c_3	0.35	0.98

component. Table 3.3 contains the memberships of the two nodes highlighted in Figure 3.5. Demand of node A is modulated (60%) by the value of demand component 1, while component 3 has a lower (35%) effect on it. On the other hand, demand of node B is completely (98%) modulated by demand component 3. Demand component 2 does not have any effect on both demands, as it is far (geographically and hydraulically) from the two example nodes. Note the similarity between binary demand components (Equation 3.4 and Figure 3.4) and hybrid demand components (Equation 3.5 and Figure 3.5).

A comparison of the calibration results between type of user-based demand patterns and pressure sensitivity-based demand components is presented in Sanz and Pérez [2014c], with better results for the latter: the uncertainty in the calibrated parameters is reduced, while the geographical distribution is useful for applications requiring parameters to be related with zones of the network.

The methodology presented next will be illustrated with the dummy networks depicted in Figure 3.6 and Figure 3.7, which represent a meshed network and a tree-like network, respectively. The simplicity of the networks will be useful to exemplify the methodology at each step.

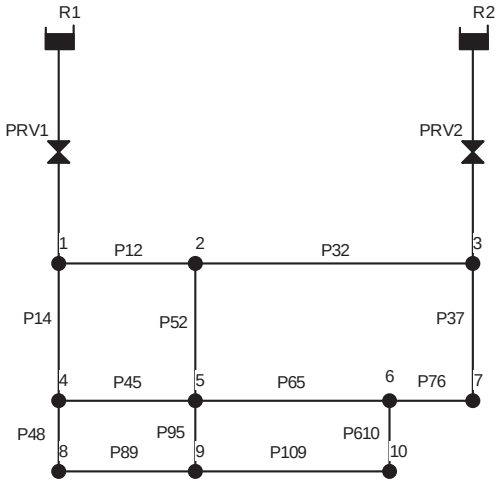

Figure 3.6: Dummy meshed network

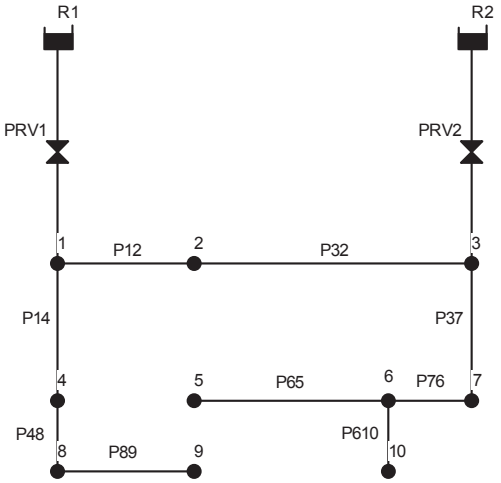

Figure 3.7: Dummy tree network

3.3 Parameter definition

As explained previously, the limited number of sensors together with the huge number of demands, requires a grouping of the parameters to make the calibration viable. In Sanz and Pérez [2014a], the authors grouped demands depending on the type of user. Although good results were obtained with synthetic data, the analysis presented in Section 3.2 encourages the use of the demand components model.

The grouping of nodal demands can be obtained from the analysis of the SVD of the system sensitivity matrix. Wiggins [1972] stated that "we can think of the eigenvectors $\mathbf{v}_i$, where $i = 1...n$, as a new parameterization of the model. These vectors represent a set of n linear combinations of the old parameters that are fixed by the observations". Similarly, we can reduce matrix $\mathbf{V}$ into $\mathbf{V_r}$, which is formed by the first n_c vectors $\mathbf{v}_i$, where n_c is the number of non-zero singular values of the sensitivity matrix. The new parameterization is obtained by defining a new parameter correction (Equation 3.6).

$$\Delta \mathbf{x}^* = \mathbf{V_r}^T \Delta \mathbf{x} \tag{3.6}$$

where $\mathbf{x}$ is the parameter vector, and $\mathbf{x}^*$ is the new parameter vector to be calibrated. In WDN very low singular values appear (as seen in Figure 3.1), thus n_c is defined in a way that all values below the n_c highest singular values are neglected. Furthermore, the consideration of very low singular values leads to increment in uncertainty (Aster et al. [2005]). The main drawback of this approach is the loss

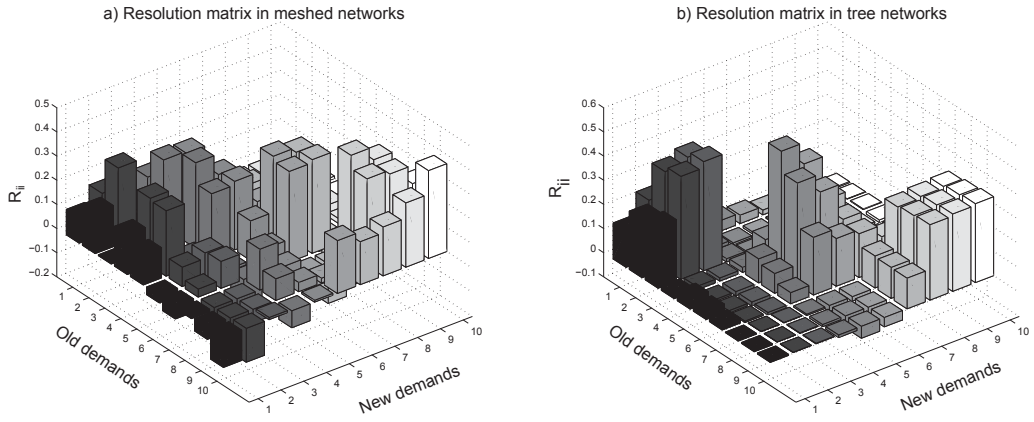

Figure 3.8: Resolution matrices defining how the new demands are represented by old demands when considering a new parameterization ($n_c = 3$) in: a) Meshed network, and b) Tree-like network

of the physical meaning of the calibrated parameters as they will be generated by a combination of the original parameters. The sensors' data will be fitted, but the calibrated parameters will not have a direct relation with the WDN.

Consequently, the objective is to define the new parameterization as a static combination of the old parameters. The resolution matrix $\mathbf{R}$:

$$\mathbf{R} = \mathbf{V}\mathbf{V}^T \tag{3.7}$$

describes how the generalized inverse solution smears out the original model $\mathbf{x}$ into a recovered model $\hat{\mathbf{x}}$ (see Appendix B). A perfect resolution is represented by the identity matrix, indicating that each parameter is perfectly resolved. When only n_c parameters corresponding to the highest n_c singular values are considered, the resolution matrix computed with $\mathbf{V_r}$ is not the identity matrix. Compact resolution appears and parameters with similar sensitivities can be identified.

In the WDN particular case, compact resolution may appear but not being easily observable in the resolution matrix, as the demand order in the sensitivity matrix $\mathbf{S}$ columns has no geographic order (in meshed networks, it is impossible to establish an order). Figure 3.8 presents the resolution matrix computed with $n_c = 3$ in: a) meshed network of Figure 3.6, and b) tree-like network of Figure 3.7. Compact resolution is visible in the tree-like network if rows and columns are correctly sorted, whereas in the meshed network compact resolution is not observable.

The identification can be performed by means of the "*delta vector generation*" process by Wiggins [1972], which aim is to generate delta vectors that form an identity matrix so that each new parameter is perfectly resolved. The process in Wiggins [1972] is adapted in this thesis to define the matrix $\mathbf{M}$ with the membership of each individual demand to each demand component. The resulting parameterization is used to calibrate groups of demands.

Pseudo-code 1 presents the whole process to generate the $\mathbf{M}$ matrix from the $\mathbf{V_r}$ reduced matrix. In lines 1-7, the delta vector generation process is performed, where the n_c vectors with the highest resolving power in the resolution matrix are obtained and normalized iteratively to generate the delta vectors. Matrix $\mathbf{V}$ is formed by orthonormal column vectors, thus the resolution matrix $\mathbf{R} = \mathbf{V}\mathbf{V}^T$ is the identity matrix, with the same number of orthonormal column vectors. As we consider less columns in $\mathbf{V}$ to form $\mathbf{V_r}$, the resolution matrix will be less like the identity matrix, and its vectors will progressively lose the orthonormal property. The process performed in lines 2-6 consists in iteratively detecting the columns from $\mathbf{R}$ that are near to be orthonormal vectors. At each iteration, the chosen column vector is divided by the square root of its diagonal value (line 4). This is done to ensure that the matrix obtained by $\mathbf{v}^*\mathbf{v}^{*T}$ will generate zeros in the column of $\mathbf{v}^*$ and almost zeros in all columns with similar directions (i.e. information) when performing the operation $\mathbf{R} = \mathbf{R} - \mathbf{v}^*\mathbf{v}^{*T}$ (line 5). Consequently, the following iteration will select a column vector with different direction (information). In lines 8-11, matrix $\mathbf{V}^*$, which is formed by the $\mathbf{v}^*$ delta vectors, is

Pseudo-code 1 Computation of nodal demands memberships to demand components

Require: $\mathbf{V_r}$, n_c, n_d
1: Compute $\mathbf{R} = \mathbf{V_r}{\mathbf{V_r}}^T$
2: **for** $z = 1 : n_c$ **do**
3: Find $j = \max(diag(\mathbf{R}))$
4: Compute $\mathbf{v}^*_z = \mathbf{R}_{(:,j)} / \sqrt{\mathbf{R}_{(j,j)}}$
5: Compute $\mathbf{R} = \mathbf{R} - \mathbf{v}^*_z \cdot \mathbf{v}^{*T}_z$
6: **end for**
7: Define $\mathbf{V}^* = [{\mathbf{v}^*}_1 \mid {\mathbf{v}^*}_2 \mid \cdots \mid {\mathbf{v}^*}_{n_c}]$
8: **for** $i = 1 : n_d$ **do**
9: Compute $\mathbf{M}_{(i,:)} = |{\mathbf{V}^*}_{(i,:)}| / \sum |{\mathbf{V}^*}_{(i,:)}|$
10: **end for**
11: **return** : $\mathbf{M}$

used to generate the $\mathbf{M}$ matrix, associating each initial parameter (consumption) to a new parameter (demand component) that produces the best resolution if n_c

demand components are considered. The normalization of the rows in $\mathbf{V}^*$ is done so that the weights can be interpreted as memberships of each nodal demand to each demand component.

Three approaches were studied in Sanz and Pérez [2015a] before reaching the final procedure: binary parameterization, positive hybrid parameterization and free hybrid parameterization.

- The first approach assigns a single demand component to each nodal demand, as seen in Equation 3.4. After executing lines 1-7 in Pseudo-code 1, each demand is associated to the demand component that has highest value in the corresponding columns of the $\mathbf{V}^*$ matrix.
- The second approach assigns a combination of demand components to each nodal demand with positive weights, exactly as presented in Pseudo-code 1.
- The free hybrid parameterization considers a combination of demand components that can include negative weights. For this approach, the normalization in line 9 of Pseudo-code 1 is ignored.

In all the proposed approaches, the solution tends to generate geographical patterns, as the topological information (incidence matrix $\mathbf{B}$) is included in the sensitivity matrix, as seen in Equation 2.13. Results obtained in Sanz and Pérez [2015a] concluded that the use of positive weights to perform the calibration of demand components gave the best results in terms of error minimization.

Figure 3.9 and Figure 3.10 depict the output of the delta vector generation process (subfigure a)), and the memberships obtained after the normalization performed in lines 8-11 of Pseudo-code 1 (subfigure b)).

Figure 3.11 and Figure 3.12 depict in each of their subfigures the memberships of each demand node to a particular demand component. The darker the color in the map, the higher the membership to the depicted demand component. The same results have already been observed in a larger network in Figure 3.5.

Pseudo-code 1 uses the sensitivity matrix computed at a particular working point. The procedure can be applied considering multiple boundary conditions to make the membership definition process more robust. However, the static topology of the network is not expected to produce significant changes in the sensitivity matrix. The application of the same process using other working points for the dummy networks generates the same memberships with only $\pm 1\%$ variations in the memberships. In the applications presented in this thesis, networks do not contain valves or pumps within the network, thus the memberships obtained are

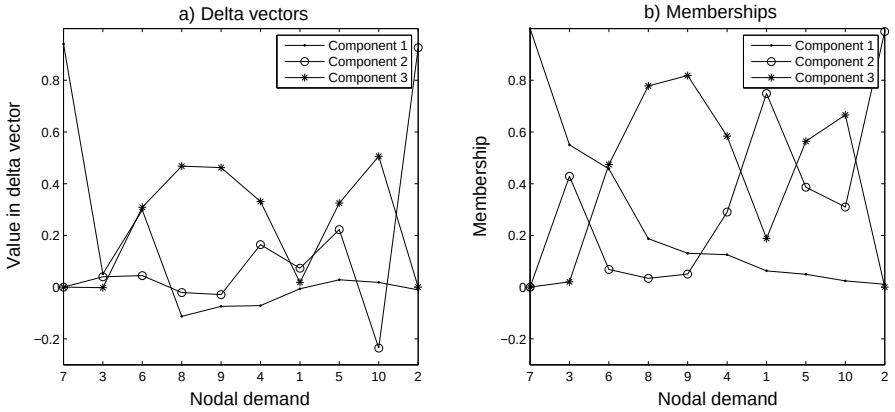

Figure 3.9: Parameterization process applied to a meshed network: a) Delta vectors, and b) Memberships of each nodal demand to each demand component

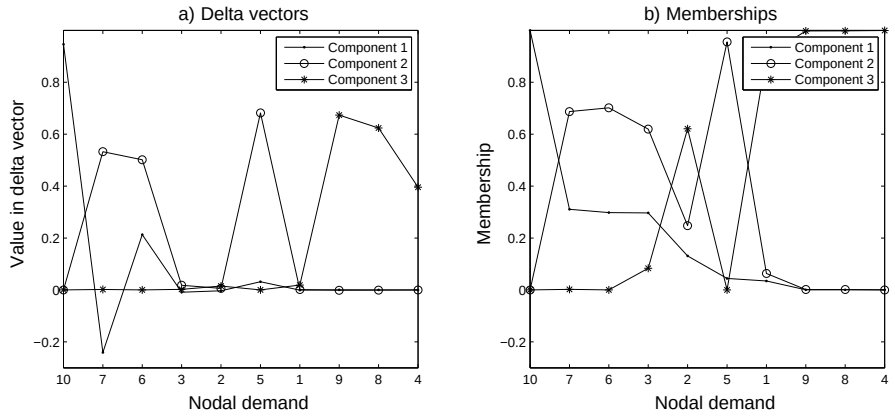

Figure 3.10: Parameterization process applied to a tree-like network: a) Delta vectors, and b) Memberships of each nodal demand to each demand component

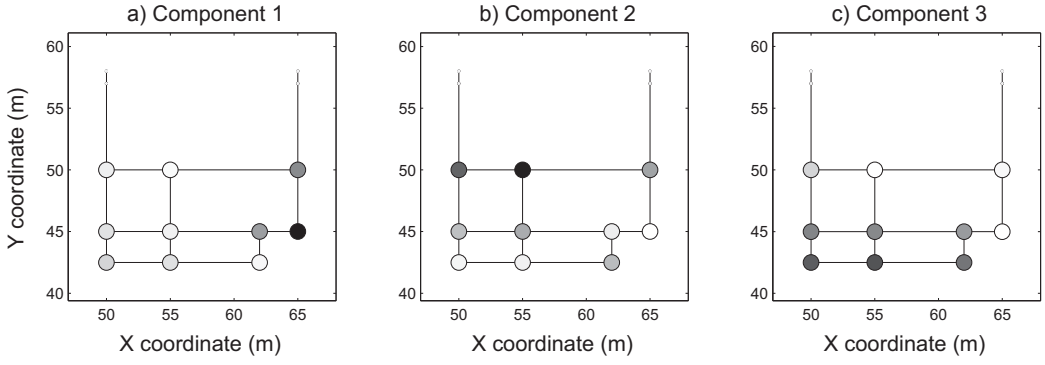

Figure 3.11: Graphical representation of the nodal memberships to demand components in a meshed network

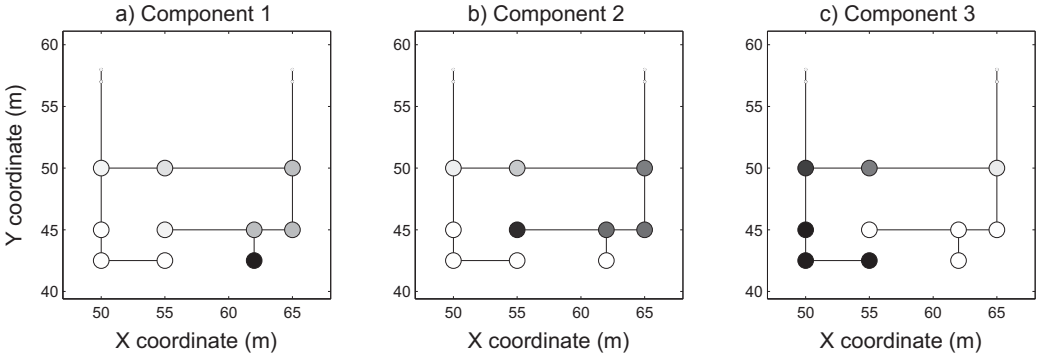

Figure 3.12: Graphical representation of the nodal memberships to demand components in a tree-like network

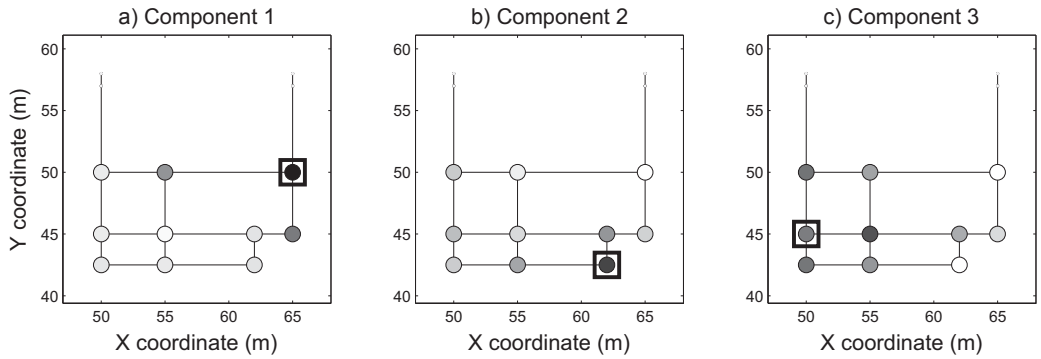

Figure 3.13: Graphical representation of the nodal memberships to demand components in a meshed network considering three installed sensors

constant during 24 hours (only really low variations due to boundary conditions). In networks with valves and pumps, the membership calculation process should be performed for each different working point, and the calibration be performed taking into account the memberships for the specific working point.

The calibration methodology requires some inner sensors to be distributed through the sampling design. In case the network already has the sensors installed, the $\mathbf{S}$ matrix introduced in Pseudo-code 1 would be a reduced sensitivity matrix $\mathbf{S_r}$ where only the rows related to the available sensors would be considered. Figure 3.13 and Figure 3.14 depict the parameterization of the two dummy networks considering that three sensors were already installed in the networks. These sensors are marked with a black square.

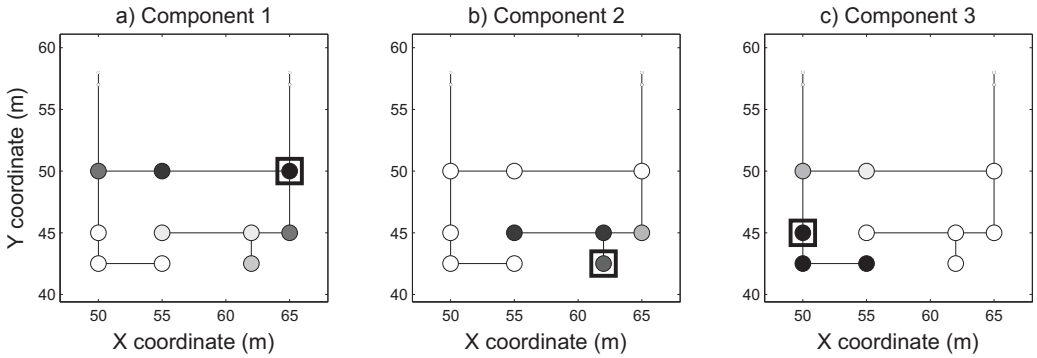

Figure 3.14: Graphical representation of the nodal memberships to demand components in a tree-like network considering three installed sensors

3.4 Sampling design

The information extracted from the network depends on the type and location of the sensors. Each new sensor represents an additional equation in the system of equations to be solved. In order to have a determined system of equations, the number of measurements (sensors) has to be at least equal to the number of parameters, guaranteeing the system identifiability in the linear approximation. The sampling design is performed after the distribution of components, selecting the n_c best sensors. The process for locating the sensors uses matrix $\mathbf{U}$ in the same way as the parameterization process uses matrix $\mathbf{V}$.

Initially, the sensitivity matrix $\mathbf{S}^*$ relating head and/or flow variations with demand components variations is computed and decomposed using the SVD. Matrix $\mathbf{U_r}$ is constructed with the first n_c columns of $\mathbf{U}$, as the information from the subsequent columns is negligible (they are multiplied by null rows of the $\mathbf{\Lambda}$ matrix). Then, the information density matrix $\mathbf{I_d}$ is computed as explained in Aster et al. [2005]:

$$\mathbf{I_d} = \mathbf{U_r}\mathbf{U_r}^T \tag{3.8}$$

describing how the generalized inverse solution smears out the original data $\mathbf{y}$ into a predicted data $\hat{\mathbf{y}}$ (see Appendix B). Since $\mathbf{I_d}$ has been constructed from n_c orthonormal vectors in $\mathbf{U_r}$, a set of n_c orthonormal vectors can be extracted from $\mathbf{I_d}$ in a way that they enhance the delta-like behavior of the $\mathbf{I_d}$ matrix (Wiggins [1972]). This "*delta-like*" vector generation process is presented in Pseudo-code 2 (lines 1-6). This process results in a set of delta-like vectors $\mathbf{u}^*$ that form matrix $\mathbf{U}^*$ (line 7). Subsequently, the rows of matrix $\mathbf{U}^*$ are normalized (line 8). This normalization prioritizes sensors that are sensitive to a unique parameter over

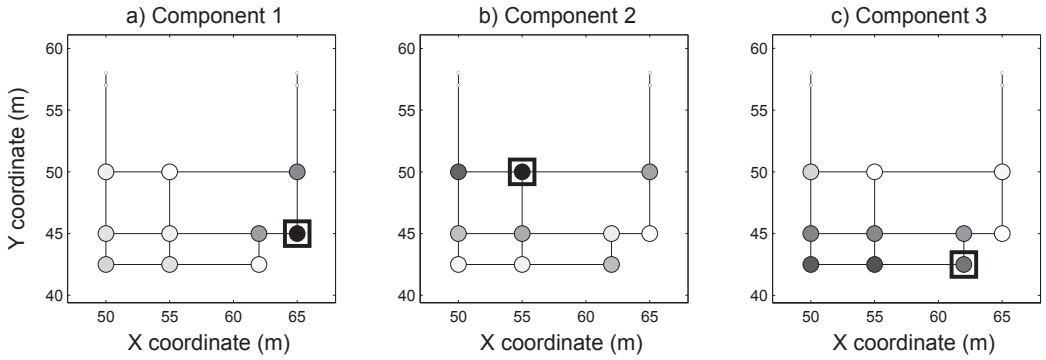

Figure 3.15: Sensor selection results applied to a meshed network with three demand components

those sensors that are sensitive to multiple parameters. Finally, the sensor with highest value in each of the n_c columns is selected as the sensor with highest information density to calibrate a particular parameter (lines 9-11). In the end, n_s=n_c sensors are selected.

Pseudo-code 2 Sensor selection process

Require: $\mathbf{U_r}$, n_s
1: Compute $\mathbf{I_d} = \mathbf{U_r}\mathbf{U_r}^T$
2: **for** $z = 1 : n_s$ **do**
3: Find $j = \max(diag(\mathbf{I_d}))$
4: Compute $\mathbf{u}^*_z = \mathbf{I_{d(:,j)}} / \sqrt{\mathbf{I_{d(j,j)}}}$
5: Compute $\mathbf{I_d} = \mathbf{I_d} - \mathbf{u}^*_z \cdot \mathbf{u}^{*T}_z$
6: **end for**
7: Define $\mathbf{U}^* = [\mathbf{u}^*{}_1 \mid \mathbf{u}^*{}_2 \mid \cdots \mid \mathbf{u}^*{}_{n_s}]$
8: Normalize rows of $\mathbf{U}^*$
9: **for** $z = 1 : n_s$ **do**
10: Find $\mathbf{s}_{(z)} = \max(\mathbf{U}^*{}_{(:,z)})$
11: **end for**
12: **return** : $\mathbf{s}$

Figure 3.15 and Figure 3.16 depict the final results of the parameterization and sampling design process. The sensor selection has been performed after the definition of the demand components (results from the previous section). In case of multiple nodes having the highest value, the sensor can be placed in any of these nodes (as it happens in Figure 3.16.c).

Pseudo-code 2 uses the sensitivity matrix computed at a particular working point. In order to make the sensor placement process more robust, the procedure can be applied k times with k different working points. This results in a maximum

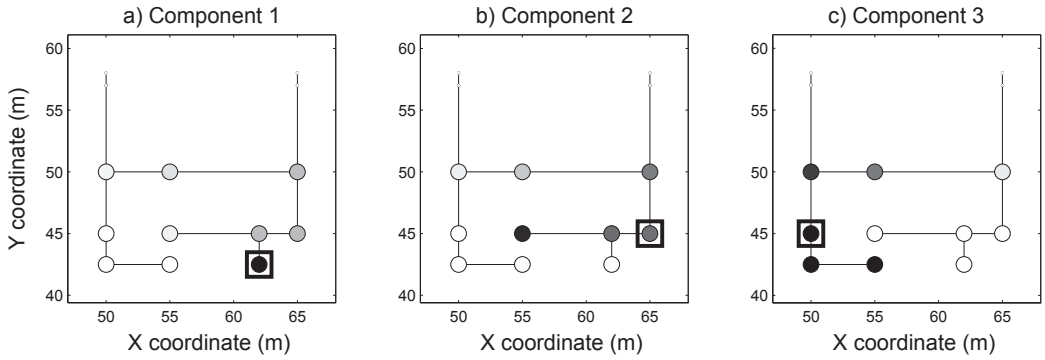

Figure 3.16: Sensor selection results applied to a tree-like network with three demand components

of $k \cdot n_s$ possible sensor locations, from which n_s sensors have to be selected. Generally, the network topology has the highest impact on the sensitivity matrix, hence the sensors chosen at each working point are placed in near (or same) locations. The repetition r_i of a particular sensor is the number of times that this sensor location has been chosen, with a maximum of $r_i = k$. Figure 3.17.a presents an example of all the possible sensors locations (and their repetitions) after applying Pseudo-code 2 with 24 different working points. The dummy networks have not been used in this case due to the low number of possible locations, which would not be helpful for the understanding of the method. The procedure to select the n_s final sensors consists of 6 steps:

1. Generate matrix $\mathbf{D_c}$ with the crossed pipe distances from each possible sensor to the others. This results in a symmetric matrix, with zeros in the diagonal.

2. Binarize the matrix, replacing distances for a "1" if the distances are lower than a predefined distance threshold d_{th}, or "0" otherwise.

3. Select the sensor with highest number of "1", i.e. the sensor with highest number of sensors within d_{th}. Figure 3.17.b shows an example of a group of sensors to be reduced.

4. For each sensor in the set, the weight $\mathbf{w}_s$ of that sensor is calculated depending on the distance to the other sensors in the set, and the repetitions of these sensors, and the sensor itself:

$$\mathbf{w}_s = \sum_{i=1}^{n_{ss}} \frac{r_i}{10^{\frac{d_{is}}{d_{max(s)}}}} \tag{3.9}$$

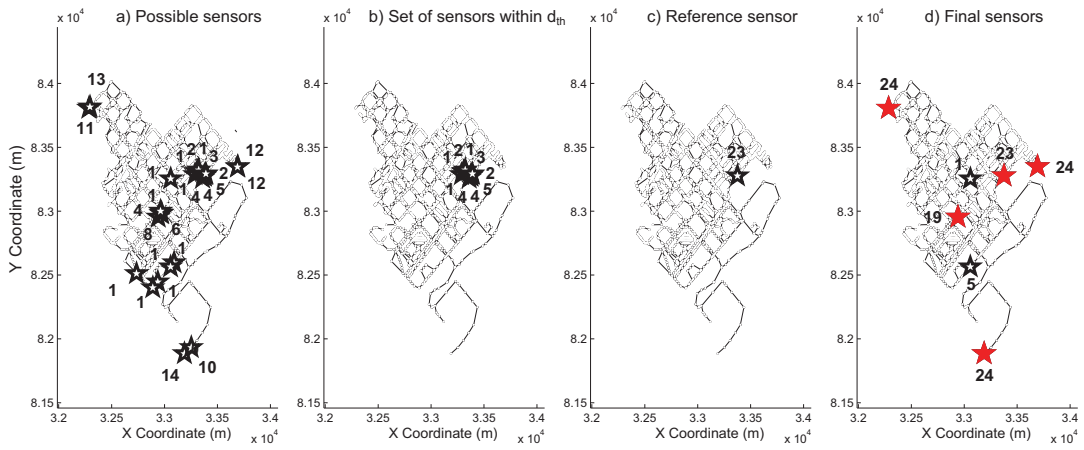

Figure 3.17: Example of the complete sampling design process: empty stars represent possible sensors, filled stars represent the selected sensors, and the numbers indicate the repetitions r_i of each sensor

where n_{ss} is the number of sensors in the current set; d_{is} is the distance between sensor s and sensor i; and $d_{max(s)}$ is the maximum distance between sensor s and all other sensors in the current set. Equation 3.9 prioritizes sensors with a high number of repetitions that are close to sensors that also have a high number of repetitions. Note that the exponent $d_{is}/d_{max(s)}$ is always in the interval $[0, 1]$, thus the denominators of the fractions are in the interval $[1, 10]$: the lower the distance d_{is}, the lower the value in the denominator.

5. The sensor in the set with highest weight is chosen as the reference sensor. All the other sensors are deleted from the possible sensors list and their number of repetitions is added to the reference sensor (Figure 3.17.c).

6. Repeat step 1 until no sets of sensors appear in the binarized distance matrix.

This process generates a number of clusters depending on the defined threshold distance d_{th}. In the end, the n_s sensors with highest repetition number from the remaining set of sensors are chosen. Figure 3.17.d shows all the possible sensors from the reduced set (stars), and the five selected sensors with highest repetitions (filled stars).

The study in Sanz and Pérez [2015b] compares the A, D, and V-optimalities from the proposed methodology solution to a significant set of possible solutions in a real network. The conclusion obtained from this study is that the solution

from the proposed approach can be little improved in terms of these optimalities when considering pressure sensors or a combination of flow and pressure sensors. Consequently, the method can be a good (and much faster) alternative to the application of GAs or exhaustive algorithms.

The sensors selection methodology presented in this section and applied to the networks throughout the thesis is based on pressure observations. The sampling design can be performed to select flow sensors if the flow/demand sensitivity matrix is considered. A comparison of using pressure and/or flow sensors for demand component calibration is presented in Sanz and Pérez [2015b].

Chapter 4

Calibration

As seen in Section 2.3, there exist multiple procedures to calibrate a water distribution network. In this chapter, an implicit method that belongs to the gradient-type procedures is presented. The method is used to calibrate the demand components of the dummy meshed network presented in Chapter 3 (Figure 3.6). The SVD is used by the methodology to compute the inverse of the sensitivity matrix and the uncertainty propagation from the used sensors to the resulting parameters.

4.1 Background theory review: the inverse problem

The objective of the calibration problem of Equation 3.1 is to find the parameter vector $\mathbf{x}$ that minimizes the errors $\boldsymbol{\varepsilon} = \mathbf{y_m} - \mathbf{y_p}(\mathbf{x})$, where $\mathbf{y_m}$ and $\mathbf{y_p}(\mathbf{x})$ are the vectors of measured and predicted values, respectively. The corrections in parameters $\Delta\mathbf{x}$ that make $\boldsymbol{\varepsilon} \to 0$ are obtained by solving the system of equations:

$$\mathbf{S} \cdot \Delta\mathbf{x} = \mathbf{y_m} - \mathbf{y_p}(\mathbf{x}) \tag{4.1}$$

where $\mathbf{S}$ is the sensitivity matrix that relates errors in predictions to corrections in the models' parameters. In non-linear problems, $\Delta\mathbf{x}$ is calculated iteratively and used to correct the parameter vector $\mathbf{x}$:

$$\mathbf{x}_{r+1} = \mathbf{x}_r + \rho\Delta\mathbf{x}_r \tag{4.2}$$

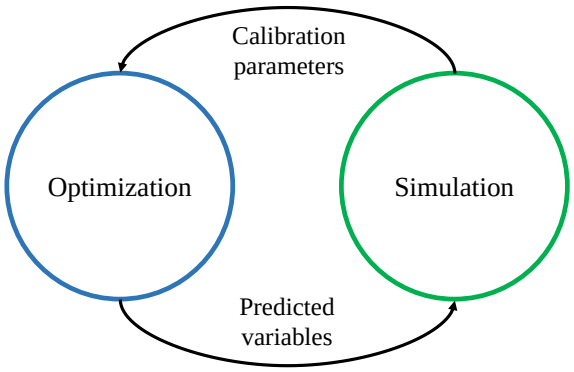

Figure 4.1: Scheme of the implicit calibration procedure

where r is the iteration number and ρ is a parameter to control the step size. The iterative scheme is continued until a termination criterion is achieved Bard [1974]. Figure 4.1 illustrates the calibration process: the optimization tool sets/updates parameters, passes them onto the simulation model, which returns back the model predicted variables. The optimization tool computes the parameters that minimize a predefined objective function.

The inverse problem may be approached by operating on both sides of Equation 3.1 with an (nxm) 'inverse' matrix $\mathbf{H}$ and letting the 'solution' or model be:

$$\hat{\mathbf{x}} \equiv \mathbf{HAx} = \mathbf{Hy} \tag{4.3}$$

The operator $\mathbf{H}$ will be a good inverse if it satisfies the following criteria Jackson [1972]:

(a) $\mathbf{AH} \approx \mathbf{I_m}$. This is a measure of how well the model fits the data, since $\mathbf{Ax} = \mathbf{y}$ if $\mathbf{AH} = \mathbf{I_m}$. The information density matrix $\mathbf{I_d}$ presented in Section 3.4 is obtained from this reasoning (see Appendix B).

(b) $\mathbf{HA} \approx \mathbf{I_n}$. This is a measure of the uniqueness of the solution, since there may exist only one solution if $\mathbf{HA} = \mathbf{I_n}$. The resolution matrix $\mathbf{R}$ presented in Section 3.3 is obtained from this reasoning (see Appendix B).

(c) The uncertainties in $\mathbf{x}$ are not too large, i.e. $var(\mathbf{x})$ is small. For statistically independent data,

$$var(\hat{\mathbf{x}}_k) = \sum_{i=1}^{m} \mathbf{H}_{ki}^2 var(\mathbf{y}_i) \tag{4.4}$$

The system in Equation 4.3 can be solved using methods such as Gaussian elimination (evendetermined), least squares method (overdetermined) or Penrose

inverse solution (underdetermined). However, none of these solution techniques can be used with rank-deficient or ill-conditioned matrices Aster et al. [2005]. The SVD is capable of solving under-, over-, even- or mixed-determined problems with no rank conditions in $\mathbf{S}$, as explained by Menke [1982]. Equation 4.1 can be solved by the SVD as:

$$\Delta \mathbf{x} = \mathbf{V} \frac{1}{\mathbf{\Lambda}} \mathbf{U}^T \boldsymbol{\varepsilon} \tag{4.5}$$

where $1/\lambda_i$ is the *ith* diagonal element of $1/\mathbf{\Lambda}$, and λ_i is the *ith* diagonal element of $\mathbf{\Lambda}$ (for $\lambda_i = 0$, the corresponding element of $1/\mathbf{\Lambda}$ is set to 0). The SVD determines the optimisation direction $\Delta \mathbf{x}$ for a problem that minimises $||\Delta \mathbf{x}||_2$ and $||\boldsymbol{\varepsilon}||_2$. A cut-off level for small λ_i is set to avoid $1/\lambda_i$ becoming too large. In this work, the cut-off level is defined at a value of $\lambda = 10^{-3}$ as suggested by Wiggins [1972] and Wasantha Lal [1995b].

The SVD was used to solve the inverse problem by Wiggins [1972] and Uhrhammer [1980] for seismographic networks, by Wasantha Lal [1995b] for unsteady river flow networks and by Cheng and He [2011] for WDN. The SVD provides a deep comprehension of the calibration problem, encouraging its adaption in the current thesis to estimate demands in WDN.

Uhrhammer [1980] and Wiggins [1972] also used the SVD matrices for the estimation of the parameter space covariance matrix to quantify the uncertainty of the calibrated model. The FOSM uncertainty quantification in Equation 2.23 can be expressed in SVD terms as

$$\mathbf{Cov}(\hat{\mathbf{x}}) = \mathbf{V} \frac{\sigma^2}{\mathbf{\Lambda}^2} \mathbf{V}^T \tag{4.6}$$

where σ^2 is the variance of the observations, represented as a scalar value if all observations have the same variance. Equation 4.6 can be generalized assuming heteroscedasticity as

$$var(\hat{\mathbf{x}}_k) = \sum_{j=1}^{n_c} \{ \sum_{i=1}^{n_c} \mathbf{V}_{ki} \lambda_i^{-1} \mathbf{V}_{ji} \}^2 var(\mathbf{y}_j) \tag{4.7}$$

If λ_i is very small, var($\hat{\mathbf{x}}_k$) will be very large. The complete formulation and derivation of the previous formulas are found in Jackson [1972]. Small λ_i values also have a direct effect on the resolution of the inverse problem. The generalized

inverse solution of Equation 4.5 can be written as:

$$\Delta \mathbf{x} = \sum_{i=1}^{n_c} \frac{\boldsymbol{\varepsilon}^T \cdot \mathbf{U}_{.,i}}{\lambda_i} \mathbf{V}_{.,i} \tag{4.8}$$

In the presence of random noise, $\mathbf{y}$ will generally have a nonzero projection onto each of the directions specified by the columns of $\mathbf{U}$ (the rows of $\mathbf{U}^T$). The presence of a very small λ_i in the denominator of Equation 4.8 can thus give a very large coefficient for the corresponding model space basis vector $\mathbf{V}_{.,i}$, and these basis vectors can thus dominate the solution. In the worst case, the generalized inverse solution is just a noise amplifier, and the answer is nonphysical and practically useless. The parameterization presented in Section 3.3 that reduces the parameters to be calibrated prevents the appearance of small λ values.

4.2 Nodal demand calibration

This thesis proposes an implicit calibration methodology that uses the SVD for solving the inverse problem. The nodal demand vector $\mathbf{d}$ is computed through the minimization of the objective function:

$$\begin{aligned} &\text{min.} && \mathcal{J}(\mathbf{d}) = \sum_{i=1}^{n_h} (w_i^h)^2 [h_i^m - h_i^p(\mathbf{d})]^2 + \sum_{j=1}^{n_f} (w_j^q)^2 [q_j^m - q_j^p(\mathbf{d})]^2 \\ &\text{subject to} && \mathcal{G}(\mathbf{h^p}, \mathbf{d}, \mathbf{r}, \mathbf{q}) = 0 \end{aligned} \tag{4.9}$$

Where $\mathcal{J}$ is the objective function; h_i^m and $h_i^p(\mathbf{d})$ are the measured and predicted heads at node i, respectively; q_j^m and $q_j^p(\mathbf{d})$ are the measured and predicted flows at pipe j, respectively; $\mathbf{w^h}$ and $\mathbf{w^q}$ are the weighting factors applied to the different terms to ensure that they are of similar magnitude and unit; n_h and n_f are the number of measured nodal heads and pipe flows, respectively; $\mathbf{r}$ is the vector of pipes' roughness; and $\mathcal{G}$ is the system of non-linear equations describing the hydraulic steady state of flows and pressures in a WDN, including mass continuity and energy conservation equations, already described in Equation 2.10 and Equation 2.9. The formulation of the generalised inverse problem in Equation 4.1 for WDN is:

$$\mathbf{S} \cdot \Delta \mathbf{d} = \boldsymbol{\varepsilon} \tag{4.10}$$

where $\mathbf{S}$ is a concatenation of the sensitivity matrices $\mathbf{A_{mh}}$ and $\mathbf{A_{mq}}$, as explained in Section 2.2; and $\boldsymbol{\varepsilon}$ is a combination of prediction errors in measured heads and

flows.

$$\begin{vmatrix} \mathbf{A_{mh}} \\ \mathbf{A_{mq}} \end{vmatrix} \Delta\mathbf{d} = \begin{vmatrix} \Delta\mathbf{h_{mh}} \\ \Delta\mathbf{q_{mq}} \end{vmatrix} \tag{4.11}$$

A constraint is added defining that the sum of consumptions can not vary, as it is assumed that the total consumed flow is known from the boundary conditions. The uncertainty in the measurement of the total network consumption is considered to be negligible.

$$\mathbf{1}\Delta\mathbf{d} = 0 \tag{4.12}$$

where $\mathbf{1}$ is a row vector of ones with length equal to the number of demands. Weights $\mathbf{W}$ are added in order to unify units.

$$\mathbf{W}\begin{vmatrix} \mathbf{A_{mh}} \\ \mathbf{A_{mq}} \\ \mathbf{1} \end{vmatrix} \Delta\mathbf{d} = \mathbf{W}\begin{vmatrix} \Delta\mathbf{h_{mh}} \\ \Delta\mathbf{q_{mq}} \\ 0 \end{vmatrix} \tag{4.13}$$

The SVD determines $\Delta\mathbf{d}$ for a problem that minimises $||\Delta\mathbf{d}||_2$ and $||\boldsymbol{\varepsilon}||_2$. As the WDN is represented by a non-linear model, $\Delta\mathbf{d}$ is calculated iteratively, and used to correct the demand vector $\mathbf{d}$:

$$\mathbf{d}_{r+1} = \mathbf{d}_r + \rho\Delta\mathbf{d}_r \tag{4.14}$$

At each iteration, the sensitivity matrices are computed using the resulting estimated demands of the previous iteration.

4.3 Demand component calibration

As noted in Section 3.3, thousands of nodal demands cannot be calibrated individually, thus nodal demands are estimated through the calibration of demand components. The demand components model in Equation 3.5 can be expressed in matrix form as:

$$\mathbf{d}(t) = \mathbf{q_{in}}(t) \cdot \mathbf{BDM} \cdot \mathbf{M} \cdot \mathbf{c}(t), \tag{4.15}$$

where $\mathbf{BDM}$ is a diagonal matrix with the average proportion of consumption of each node over the total consumption, computed from billing; $\mathbf{M}$ contains the memberships of every node to each demand component; and $\mathbf{c}(t)$ is the vector of demand components values. Considering that errors in nodal demands $\Delta\mathbf{d}$ are

provoked by errors in demand components $\Delta\mathbf{c}$, Equation 4.10 can be written as:

$$\mathbf{q_{in}}(t) \cdot \mathbf{S} \cdot \mathbf{BDM} \cdot \mathbf{M} \cdot \Delta\mathbf{c}(t) = \boldsymbol{\varepsilon}(t). \tag{4.16}$$

The SVD determines $\Delta\mathbf{c}$ for a problem that minimizes both $||\Delta\mathbf{c}||_2$ and $||\boldsymbol{\varepsilon}||_2$. Solving iteratively Equation 4.16 leads to a set of demand components values for time t.

4.4 Weights selection

The selection of the weights is an important factor when working with different types of units. Cheng and He [2011] used the coefficients in matrix A to transform the water head errors to flow errors by multiplying the diagonal element $\mathbf{A}_{ii}$ in both sides of the rows in Equation 4.13 that correspond to heads. $\mathbf{A}_{ii} \cdot \Delta\mathbf{h}_i$ represents the increase of water flowing into the node if node head increase by $\Delta\mathbf{h}_i$. However, the existence of very low head losses generates high values in these weights, leading to errors in the calibration process. Consequently, the weights in this thesis have been defined so that the pressure and flow errors are converted into relative errors. Consequently, the optimization process generates the demand components values that minimize the norm of the relative errors, as seen in Equation 4.17.

$$\mathcal{J}(\mathbf{d}) = \sum_{i=1}^{n_h} [\frac{h_i^m - h_i^p(\mathbf{d})}{h_i^m}]^2 + \sum_{j=1}^{n_f} [\frac{q_j^m - q_j^p(\mathbf{d})}{q_j^m}]^2 \tag{4.17}$$

4.5 Reducing the uncertainty

The stochastic nature of demands and the noise in measurements cause the calibrated demand components to have high uncertainty. Two actions are proposed to reduce it: oversampling and inclusion of extra data.

Oversampling

The reduction of the sensors' noise can be achieved by filtering several measurements along a period of time, reducing the standard deviation of the averaged measurement (Equation 4.18) while increasing its resolution (Pandya and Gupta [2014]).

$$\sigma_{filtered} = \frac{1}{\sqrt{n_k}} \cdot \sigma_{measurement} \tag{4.18}$$

where n_k is the number of measurements taken during the same time period. In this thesis the oversampling is used to filter the measurements of each sensor throughout an hour, so that a single demand component value represents the demand behavior during that hour. The simple average value is considered, as all samples within the hour have the same importance.

Including extra data

The inherent noise in demands that affects the calibrated values can be reduced if data from multiple samples with the same boundary conditions and same expected demand behaviors are used. The system becomes overdetermined, and the solution minimizes the error of all samples simultaneously. The system presented in Equation 4.16 is extended including the extra data samples, as shown in Equation 4.19.

$$\mathbf{W}\begin{vmatrix} \mathbf{A_{mh}}(k_1)\cdot\mathbf{q_{in}}(k_1) \\ \vdots \\ \mathbf{A_{mh}}(k_n)\cdot\mathbf{q_{in}}(k_n) \\ \mathbf{A_{mq}}(k_1)\cdot\mathbf{q_{in}}(k_1) \\ \vdots \\ \mathbf{A_{mq}}(k_n)\cdot\mathbf{q_{in}}(k_n) \\ \mathbf{1} \end{vmatrix} \cdot \mathbf{BDM}\cdot\mathbf{M}\cdot\Delta\mathbf{c}(t\rightarrow k_1\ldots k_n) = \mathbf{W}\begin{vmatrix} \Delta\mathbf{h_{mh}}(k_1) \\ \vdots \\ \Delta\mathbf{h_{mh}}(k_n) \\ \Delta\mathbf{q_{mq}}(k_1) \\ \vdots \\ \Delta\mathbf{q_{mq}}(k_n) \\ 0 \end{vmatrix} \tag{4.19}$$

The inclusion of the extra samples does not increase the rank of the sensitivity matrix. However, the condition number decreases Aster et al. [2005], reducing the propagation of noise from measurements to demand components (Equation 4.6)(Pérez [2003]).

4.6 Online application

The online demand components calibration process can be explained through the scheme in Figure 4.2. Sensors in the network take periodic measurements that are stored in the database through the SCADA system. All data are analyzed to detect missing data, spurious measurements, trends, etc. At a particular day d and hour h, the calibration process takes from the database a set of measurements and boundary conditions, corresponding to that day d and hour h. The number of measurements taken by each sensor within that hour depends on the sensor

sampling time: the lower the sampling time, the higher the number of measurements taken in an hour. The multiple samples from the same hour at each sensor are filtered to reduce the sensors' noise and demand uncertainty effect.

Inside the calibration process, a simulation is run for each set of boundary conditions available. The predicted measurements obtained from the simulations are filtered, so that they can be compared with the real filtered measurements. The resulting prediction error is minimized by the correction of the demand components values.

Extra data from previous days at the same hour h can be used in the calibration process. The calibration horizon H_c defines the number of days used to calibrate a single set of demand components values. The method assumes that the boundary conditions and demand behaviors from the H_c days used are similar. When the calibration process finishes, the resulting demand components values can be used in any model-based application needing a calibrated hydraulic model such as leakage localization Goulet et al. [2013], quality modeling Nejjari et al. [2011] or leak detection based on demand components analysis Sanz et al. [2015].

4.7 Exemplification

The methodology presented in this chapter is applied to the dummy meshed network in Figure 3.6. The memberships of nodes to demand components and the selected sensors have already been defined in Chapter 3 (Figure 3.11 and Figure 3.15). The calibration process will calibrate three demand components that minimize the error in the predicted pressures at the three selected sensors. The total consumption of the network is assumed to be known, but not the distribution of this consumption among the inputs. The calibration process is performed during 24 hours, considering a calibration horizon $H_c = 1$ (only data from the current day is used to calibrate the components), and one sample per hour (there will be no filtering inside each hour).

Figure 4.3 presents the evolution of the pressure prediction RMSE and how this error makes the demand components vary at each iteration. It can be seen that the lower the RMSE, the lower the variation of the demand components values.

Once the termination criterion is achieved (maximum number of iterations, or number of iterations with no significant reduction in the RMSE), the calibration

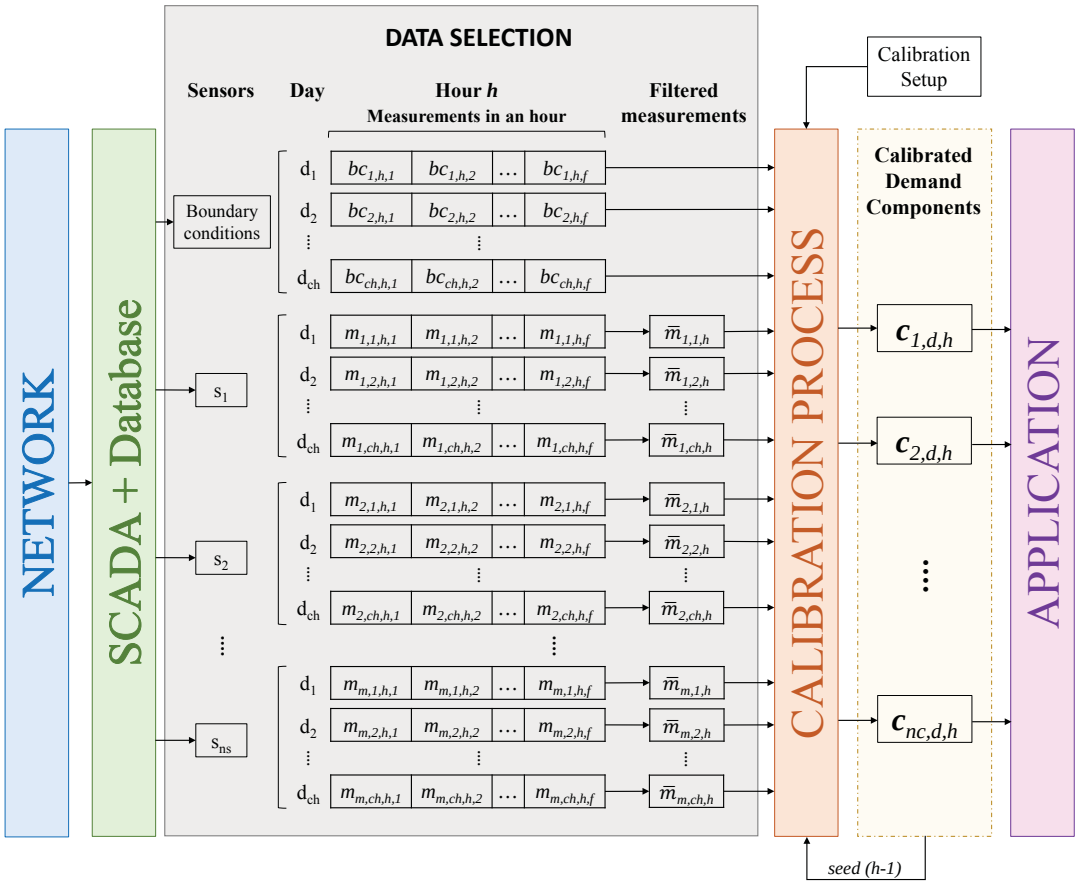

Figure 4.2: Scheme of the whole calibration process. Data from the network is obtained and validated via SCADA and stored into the database. At each hour, a set of filtered measurements and boundary conditions is introduced into the calibration process. Each sensor takes f measurements per hour. Additional measurements from $H_c - 1$ previous days can also be used to enhance the calibration. The calibration process estimates a set of demand components values that minimize the error in the filtered measurements. Applications can make use of the calibrated components to improve the hydraulic model

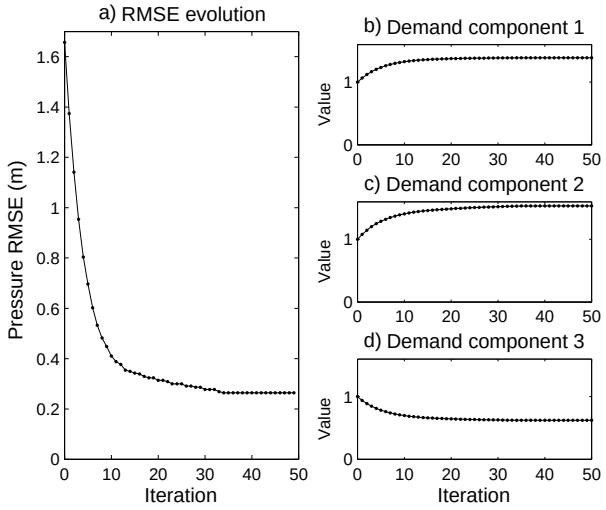

Figure 4.3: Pressure prediction RMSE and demand components evolution during the iterative calibration applied to the dummy meshed network for h=15

Table 4.1: Dummy meshed network pressure and flow prediction RMSE using the basic demand model and the demand components model. The percentage of improvement with respect to the basic demand model is also presented

	Basic Model RMSE	**Demand Components Model RMSE**	**Improvement**
Pressure (m)	1.56	0.28	82%
Flow (l/s)	0.5	0.43	14%

for that specific hour is finished. The calibrated demand components values are used as the starting point (seed) for the next hour. Figure 4.4 depicts the calibrated demand components during 24 hours, and the 95% confidence intervals (CI) computed as explained in Section 2.5 and Equation 4.6.

Table 4.1 collects the pressure prediction root mean squared error (RMSE) (first row), and the flow prediction RMSE (second row), for each demand model used. The percentage of error improvement has been also computed for each type of measurement, which represents the error improvement when using the demand components model instead of the basic demand model.

The calibration process can be repeated, but this time adding the information from the flow distribution at the network inputs. Table 4.2 collects the same information as before, for the calibration that includes the flow measurements. It can be seen that the pressure RMSE improvement is lower compared to the previous results, but the flow RMSE improvement is much better.

Finally, if we consider data from 5 days ($H_c = 5$) with similar expected de-

Table 4.2: Dummy meshed network pressure and flow prediction RMSE using the basic demand model and the demand components model. The percentage of improvement of the demand components model with respect to the basic demand model is also presented. These results have been obtained with the calibration that includes the flows measured at the network inputs

	Basic Model	**Demand Components Model**	
	RMSE	**RMSE**	**Improvement**
Pressure (m)	1.56	0.55	64.7%
Flow (l/s)	0.5	0.11	78%

mand behaviors to calibrate simultaneously the demand components, the resulting uncertainty will be reduced, as explained in Section 4.5. Figure 4.5 depicts the calibrated demand components during 24 hours that minimize simultaneously the hourly pressure prediction error from five days. It can be seen that the confidence intervals when considering $H_c = 5$ are narrower than the ones obtained when using $H_c = 1$.

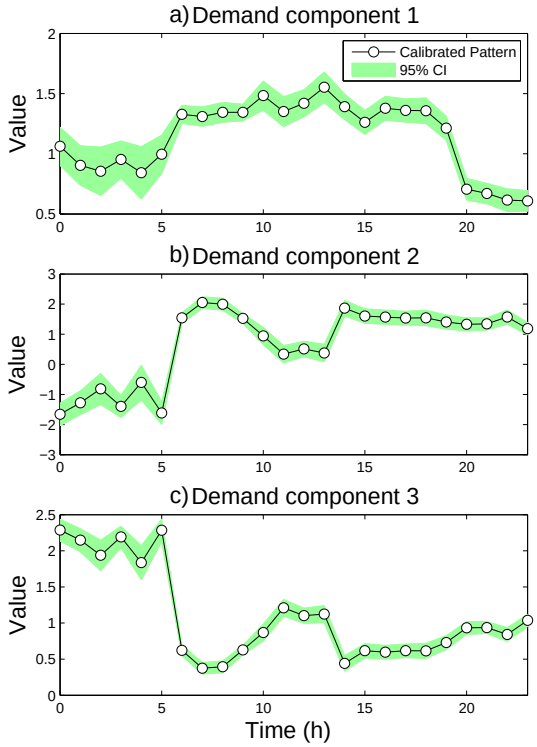

Figure 4.4: Dummy meshed network calibrated demand components with 95% confidence intervals (CI) (green boundaries) using $H_c = 1$

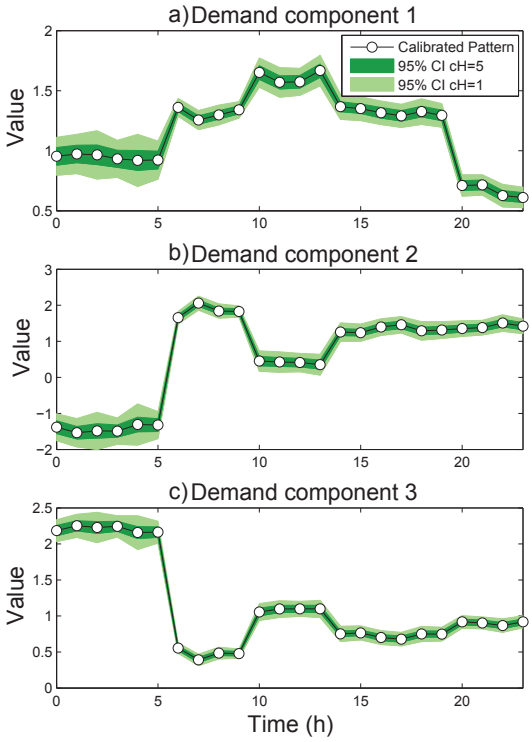

Figure 4.5: Dummy meshed network calibrated demand components with 95% confidence intervals (CI) for $H_c = 5$ (dark green) and $H_c = 1$ (light green)

Chapter 5

Leak detection and localization

The model-based leak detection and localization methodologies reviewed in Section 2.7 can use the online calibrated water distribution network model. However, these methodologies do not consider the evolution of demands in the real system. This evolution should be taken into account because demands are parameters that change continuously and leakages may be masked with their evolution.

A leak detection and localization approach coupled with the online calibration method presented in Chapter 4 is proposed. The main objective is to diagnose if the updates in the demand model during the continuous calibration correspond to the evolution of demands or to leakage. If leakage is detected, the geographical distribution of demand components enables identification of a particular zone of the network where leakage is most likely located. This leakage can be a burst or any event that induce similar abnormal pressure/flow variations at the DMA level.

The methodology proposed is used to analyze the calibrated demand components of the dummy meshed network presented in Chapter 3 (Figure 3.6), and calibrated in Chapter 4, in order to detect abnormal behaviors produced by leakage, while ignoring small variations due to demand evolution.

5.1 Leak detection and localization structure

Figure 5.1 presents the structure of the coupled calibration and leak detection and localization methodologies. Measurements taken from the real network are introduced via the SCADA system, where a validation process is performed first. The **calibration process** estimates every hour the set of current demand components $\mathbf{c}^c$ that minimize the errors in model predictions. This set of calibrated

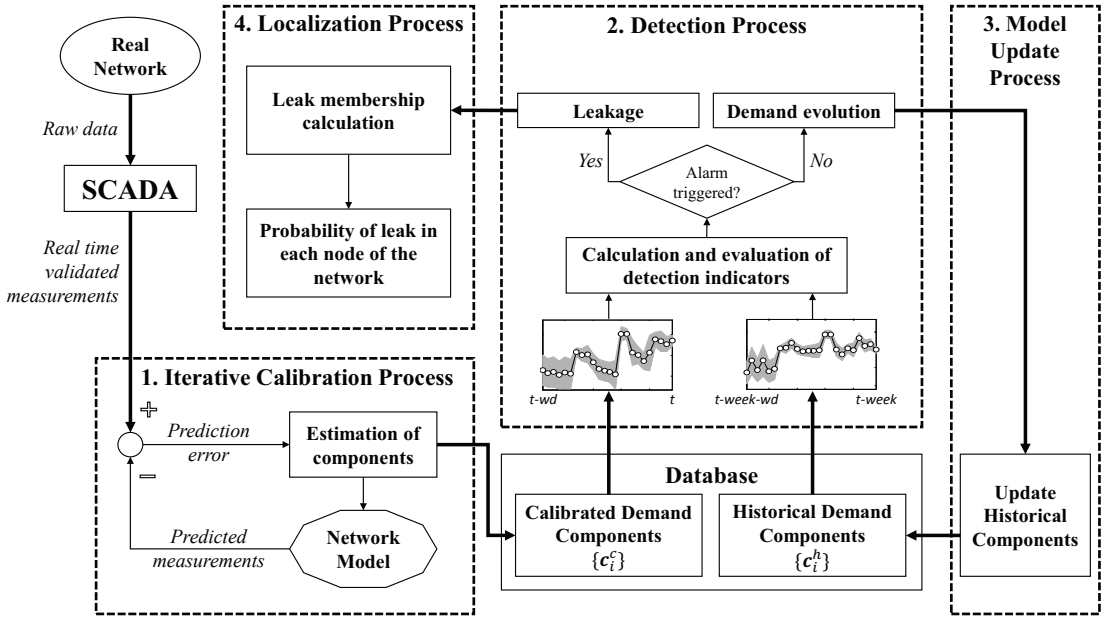

Figure 5.1: Scheme of the calibration process coupled with the model update/leakage detection and localization processes

demand components is stored into a database, where it is concatenated to previous hours and days. Simultaneously, the **detection process** compares the sets of calibrated and historical demand components. Assuming that consumers' habits do not change significantly from one week to another, the value $\mathbf{c}_i^c$ of the calibrated demand component i is expected to be similar to the corresponding value in the previous week $\mathbf{c}_i^h$ (historical component). At time t, the last wd values of each component $\mathbf{c}_i^c$ are compared with the same time window of $\mathbf{c}_i^h$ using detection indicators, where wd is the number of samples to be compared (e.g. if $wd = 24$, 24 hours of $\mathbf{c}_i^c$ will be compared with the same 24 hours of $\mathbf{c}_i^h$). If detection indicators do not trigger the detection alarm, the state of the network is classified as non-faulty, and the historical demand components values are updated with the currently calibrated ones (**model update process**). The new demand components include slight changes in demands due to the evolution of the system. On the contrary, if the detection alarm is triggered, the leak **localization process** starts. The week-to-week comparison is useful not only for the similarity of the compared days, but also to avoid false alarms from progressive changes due to seasonal habits in population.

5.2 Detection process

The detection process consists in the comparison of two sets of demand components values using detection indicators. First, the detection indicators have to be selected. Then, the criterion to decide whether the detection indicators are detecting and anomaly or not is defined by means of the computation of thresholds for each detection indicator.

5.2.1 Detection indicators

Six detection indicators are defined to evaluate the similarity or dissimilarity between calibrated and historical demand components: *Pearson correlation, conditional overlapping, L1-norm, relative increment in mean component values* and *consumption,* and *relative residual coefficient.* A description of each indicator is listed next:

- The *Pearson correlation* is a measure of the linear dependence between the two components $\mathbf{c}_i^c$ and $\mathbf{c}_i^h$.

$$\boldsymbol{\rho}_i(t) = \frac{\sum_{k=t-wd+1}^{t}[(\mathbf{c}_i^c(k) - \bar{c_i^c})(\mathbf{c}_i^h(k) - \bar{c_i^h})]}{\sqrt{\sum_{k=t-wd+1}^{t}(\mathbf{c}_i^c(k) - \bar{c_i^c})^2 \cdot \sum_{k=t-wd+1}^{t}(\mathbf{c}_i^h(k) - \bar{c_i^h})^2}} \tag{5.1}$$

  where $\mathbf{c}_i^c$ comprises times from $t - wd + 1$ to t, and $\mathbf{c}_i^h$ comprises the same times but corresponding to the previous week; and t is a specific point in time where calibrated components are available. Correlations close to 1 indicate a high similarity between components.

- The *overlapping coefficient* measures the overlap between two discrete or continuous probability density functions (pdf).

$$\mathbf{o}_i(k) = \int_{-\infty}^{\infty} \min(f_i(x), g_i(x))\ dx \tag{5.2}$$

  where $f_i(x)$ is the pdf of the current calibrated component at sample k; and $g_i(x)$ is the pdf of the historical component at the same sample of the previous week. $f_i(x)$ and $g_i(x)$ are assumed to be Gaussian distributions from the calibrated mean demand components values and standard deviations. The mean overlapping $\bar{\mathbf{o}}_i$ during a time window is calculated as seen

in Equation 5.3.

$$\bar{\mathbf{o}}_i(t) = \frac{1}{wd} \sum_{k=t-wd+1}^{t} \mathbf{o}_i(k) \tag{5.3}$$

A 100% overlap is obtained with equal probability distributions. As the pdfs become different, the overlapping decreases. A new indicator called *conditional overlapping* coefficient can be defined considering only the reduction of overlapping coefficients due to positive component changes (increase in consumed water).

$$\mathbf{co}_i(t) = \begin{cases} \bar{\mathbf{o}}_i(t) & \bar{c}_i^c > \bar{c}_i^h \\ 100\% & \text{otherwise} \end{cases} \tag{5.4}$$

- Norms are functions that assign a strictly positive length or size to a vector in a vector space, other than the zero vector.

$$||\mathbf{c}_i^c - \mathbf{c}_i^h||_p(t) = \sqrt[p]{\sum_{k=t-wd+1}^{wd} |\mathbf{c}_i^c(k) - \mathbf{c}_i^h(k)|^p} \tag{5.5}$$

Only the *L1-norm* ($p = 1$) is considered.

- The *relative increment in mean component values* $\Delta\mathbf{c}_i$ indicates the percentage of relative increment between the current values (averaged through a defined time) and the historical ones (also averaged).

$$\Delta\mathbf{c}_i(t) = 100 \cdot \frac{\bar{c}_i^c - \bar{c}_i^h}{\bar{c}_i^h} \tag{5.6}$$

where the means have been computed during a time interval wd.

- The *relative increment in mean component consumption* $\Delta\mathbf{c^d}_i$ indicates the percentage of relative increment between the current consumption (averaged through a defined time) and the historical one (also averaged). This indicator is similar to the previous one, but the components' consumptions in l/s are used instead of the dimensionless values.

$$\Delta\mathbf{c^d}_i(t) = 100 \cdot \frac{\sum_{k=t-wd+1}^{wd}(\mathbf{c}_i^c(k) \cdot \mathbf{q_{in}}^c(k)) - \sum_{k=t-wd+1}^{wd}(\mathbf{c}_i^h(k) \cdot \mathbf{q_{in}}^h(k))}{\sum_{k=t-wd+1}^{wd}(\mathbf{c}_i^h(k) \cdot \mathbf{q_{in}}^h(k))} \tag{5.7}$$

where superscripts ^{c} and ^{h} in $\mathbf{q_{in}}$ refer to current and historical total inflow, respectively.

- The *relative residual coefficient* gives a measure about the relative variation between two probability distributions considering the 95% confidence intervals.

$$\mathbf{rRes}_i(t) = \frac{100}{wd} \sum_{k=t-wd+1}^{t} \frac{(\mathbf{c}_i^c(k) - 1.96\sigma_{\mathbf{c}_i^c}(k)) - (\mathbf{c}_i^h(k) + 1.96\sigma_{\mathbf{c}_i^h}(k))}{|\mathbf{c}_i^h(k) + 1.96\sigma_{\mathbf{c}_i^h}(k)|} \tag{5.8}$$

  This measure only gives positive values when the current component lower bound is higher than the 95% upper bound of the historical component.

5.2.2 Setting of thresholds

The presented detection indicators evaluate the variation in demand components by comparing the current components' values with the previous week ones. As the variations become higher, the probability of having an anomaly in the network increases. Variations in demand components have different effects on detection indicators; e.g. the L1-norm is sensitive to changes in the component average value, whereas the conditional overlapping only considers positive changes in it. Therefore, the six indicators are combined to obtain a more robust detection.

Each detection indicator gives a score to each demand component depending on its variation. The sum of scores is then used to decide if the component has an anomaly or not. The scores given by the detection indicators depend on thresholds. The definition of a unique threshold for each indicator may produce poor leakage detection or excessive false alarms. Instead, two thresholds are defined for each indicator, giving 1 or 2 score points when overtaking the first and second threshold, respectively. Detection indicators' thresholds are defined separately, but shared by all demand components.

The thresholds values are determined through a training process when no leakage is present in the network. The mean and standard deviation of each detection indicator are computed during the non-faulty scenario. Then, the low and high detection threshold are defined at P_{80} and P_{95} respectively, where P refers to the percentile, for the worst component in each indicator. The worst case is used to avoid false alarms. Finally, the global threshold (sum of individual scores) is set so that the total sum of the non-faulty indicators is under this value. The thresholds setting proposed is performed in a way that if the network remains in the same state, the probability of data falling outside thresholds is 20% for the lower detection threshold and 5% for the higher one, for the worst component in

each indicator.

In the end, we have a system that triggers the alarm in a particular demand component if the total score for that component is higher than the global threshold.

5.2.3 Effect of undetected anomalies

Setting the thresholds for the leak detection and localization process is assumed to be done over a non-faulty state of the network. However, different types of errors or anomalies can exist both in the model or network, like undetected bursts, existing background leakages, unknown status valves (Walski et al. [2014]), or bad estimated roughness, among others. The presence of these anomalies can be treated depending on when the anomaly has appeared without being aware of it:

1. Before setting the thresholds: The undetected anomaly will hinder the best demand adjustment. Nevertheless, this anomaly will be incorporated into the calibrated demand components model. Consequently, the methodology will be able to detect new bursts that cause a change in the components from that moment on.

2. After setting the thresholds: The currently calibrated demand components will accommodate their values to adapt to the new network pressures, provoking a change compared to the historic demand components. Future studies will analyze this scenario to observe if the methodology is able to detect and locate the non-burst anomalies. These events may induce similar pressure-flow variations in the network as the ones produced by bursts.

5.3 Localization process

The detection indicators presented in previous sections are able not only to detect the leakage, but also to classify it in a determined demand component, which is associated to a specific zone of the network, thanks to the sensitivity-based memberships defined in Section 3.3. This section presents two methods (*direct method* and *leak membership method*) to interpret the geographical information contained in the nodes' memberships and locate the detected leak.

The *direct method* locates the leak depending on the membership of each node to the abnormal demand component. The higher the membership of a node to the abnormal component, the higher the probability of leak occurring close to

that node. The geographical distribution of demand components will indicate a particular zone in the network with high probability to contain the leak.

The *leak membership method* consists in calculating the theoretical leak memberships to demand components. When leakage is present, pressures decrease due to the increasing flow. Consequently, the calibration process modifies the demand components values to adapt the model to the new pressures. Therefore, all components suffer higher or lower variations that can be attributed to the leak. The demand components consumptions variations $\Delta\mathbf{cd}$ at a particular time t can be computed as

$$\Delta\mathbf{cd}_i(t) = \frac{\sum_{k=t-wd+1}^{wd}(\mathbf{c}_i^c(k)\cdot\mathbf{q_{in}}^c(k)) - (\mathbf{c}_i^h(k)\cdot\mathbf{q_{in}}^h(k))}{wd} \quad i = 1...n_c \tag{5.9}$$

which represents the absolute variations in demand components consumptions (Equation 5.7 presented the relative percentage of variations). Subsequently, the theoretical leak memberships are obtained as

$$\mathbf{l_{memb}} = \frac{|\Delta\mathbf{cd}|}{\max(|\Delta\mathbf{cd}|)} \tag{5.10}$$

Subsequently, the leak memberships are compared with the ones from all network nodes using the Pearson correlation. The higher the correlation in a node, the higher the probability of that node to contain the leak.

5.4 Exemplification

The methodology presented in this chapter is applied to the dummy meshed network in Figure 3.6. The memberships of nodes to demand components and the selected sensors have already been defined in Chapter 3 (Figure 3.11 and Figure 3.15). The calibration process will calibrate three demand components that minimize the error in the predicted pressures at the three selected sensors. The total consumption of the network is assumed to be known, but not the distribution of this consumption among the inputs. The calibration process is performed during 5 days (120 hours), considering a calibration horizon $H_c = 1$ (only data from the current day is used to calibrate the components), and one sample per hour (there will be no filtering inside each hour). At the beginning of the fifth day (hour 97) a leak appears at node '7' (see Figure 3.6). The leak is simulated using a coefficient emitter $C_e = 0.7\ l\cdot s^{-1}\cdot m^{-0.5}$, which represents

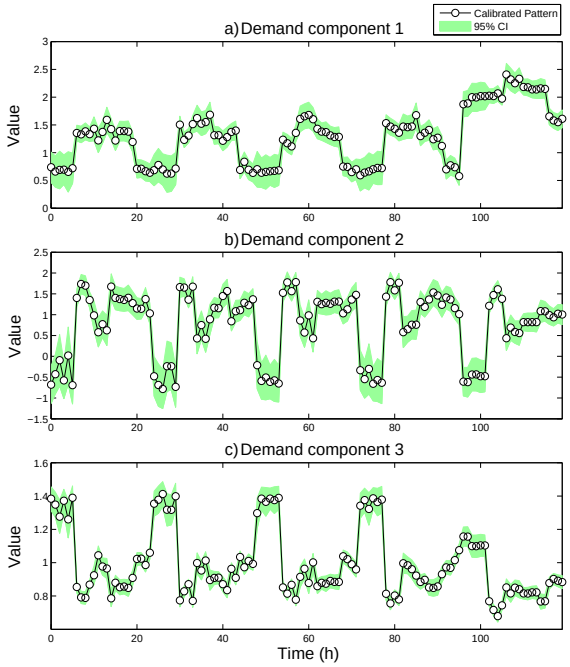

Figure 5.2: Dummy meshed network calibrated demand components with 95% confidence intervals (CI) during 5 days. A leak with an average daily water discharge appears at node '7' at h=97

an average daily water discharge of 4.25 l/s in the current network (10% of the average daily consumption in the whole network). The average pressure in the network is 36.8 m.

Data from the first day to the third day are considered as non-faulty data, and are used to compute the detection indicators' thresholds and the global threshold. Data from the fourth day to the fifth day are used to test the methodology. Figure 5.2 presents the three calibrated demand components during the five days. The leak appearance is easily seen in demand component $\mathbf{c}_1$ at hour 97, where the component value increases compared to the previous days.

5.4.1 Setting of thresholds

Data from the first three days are used to set the detection indicators thresholds. As explained in subsection 5.2.2, the lower threshold is defined so that the 80% of the detection indicators values obtained during the non-faulty scenario do not overcome this threshold, whereas the higher threshold is defined in a way that the 95% of the detection indicator values obtained during the non-faulty scenario do not overcome it. All the detection indicators are computed with $wd = 12h$

except for the correlation indicator, which is computed with a window of 24h.

Figure 5.3 depicts, for the three first days of demand components available, the values obtained for each detection indicator, and the two detection thresholds defined for each of them. The values start at hour 48h because two days of data are required to start computing the correlation indicator. The highest score obtained during the non-faulty scenario corresponds to demand component $\mathbf{c}_2$ (green line with circles) at h=59, with a value of 4 (2 points of score from the relative mean value increase plus 2 points of score from the relative mean consumption increase). Consequently, the global threshold is set at a value of 5 to avoid false alarms during the non-faulty scenario.

5.4.2 Detection

Once the thresholds have been set, data from the fourth and fifth days are used to evaluate the detection methodology. At each hour, the detection indicators are computed. If any of the computed detection indicators overtakes a threshold, the indicator gets a score depending on the threshold overtaken (1 point of score for the lower threshold, and 2 points of score for the higher threshold). Figure 5.4 depicts the values for each detection indicator during the whole scenario (three days from the training stage and two days used for the evaluation). The cyan dash-dotted vertical line indicates the time when the leak starts (h=97). It can be seen that all the indicators computed for demand component $\mathbf{c}_1$ rapidly overtake both thresholds. The slowest indicator is the conditional overlapping, where the higher threshold is overtaken 8 hours after the leak appearance.

Figure 5.5 presents the scores of each of the detection indicators presented in Figure 5.4 for the demand component $\mathbf{c}_1$. This demand component is the only component to trigger the global detection alarm, one sample after the leak appearance (signaled in the figure with a cyan dash-dotted line).

5.4.3 Localization

Once the leak has been detected, the *direct approach* gives an immediate leak localization by looking at the memberships of each network node to the demand component that has triggered the alarm. In this case, the memberships of all nodes to demand component $\mathbf{c}_1$ can be observed in Figure 3.9. The nodes with highest membership to this component are nodes '3' (56%), '6' (48%) and '7' (100%). Consequently, the indications for the water utility would be to seek for

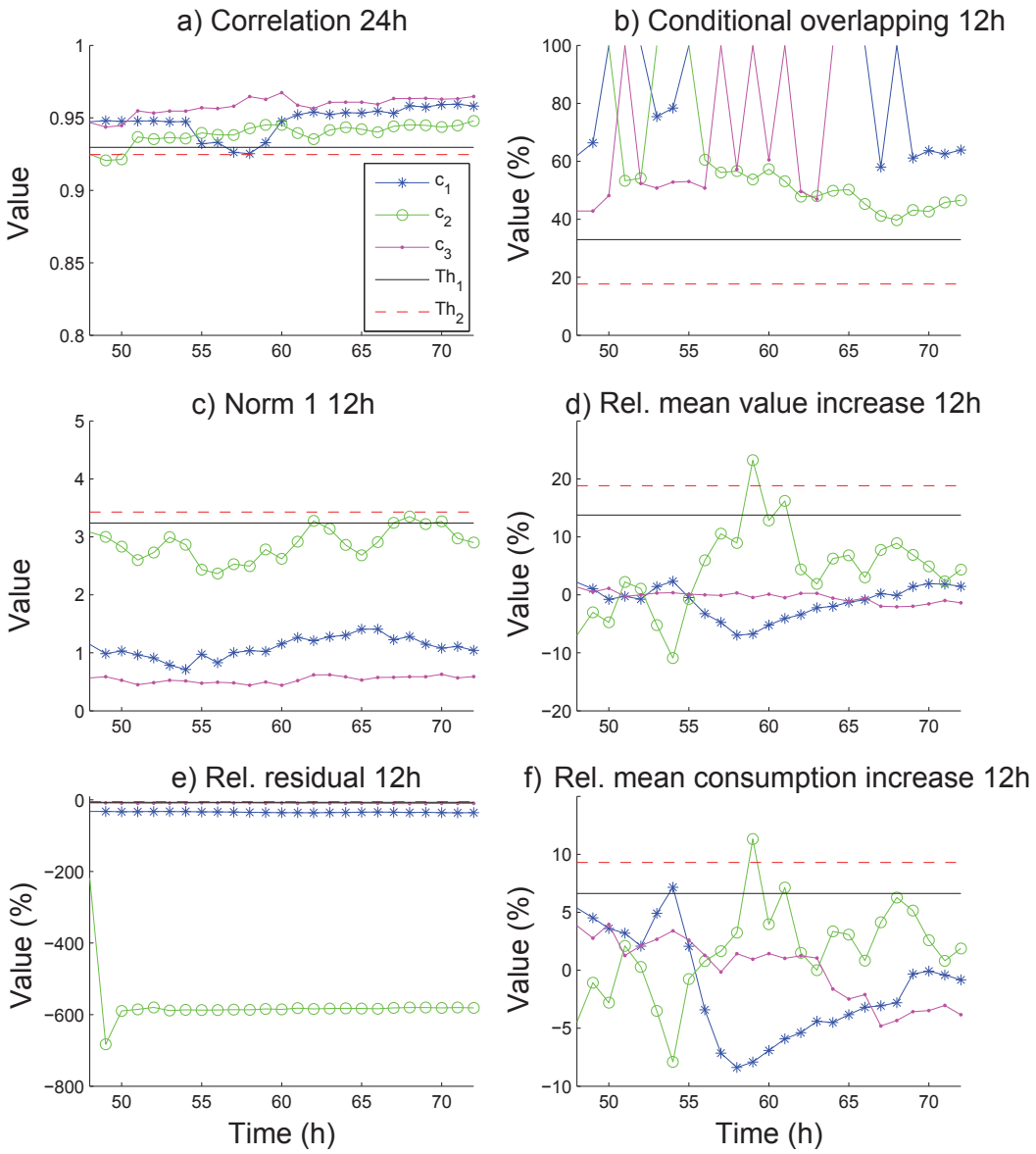

Figure 5.3: Dummy meshed network detection indicators values and set thresholds using the first three days of data (training phase)

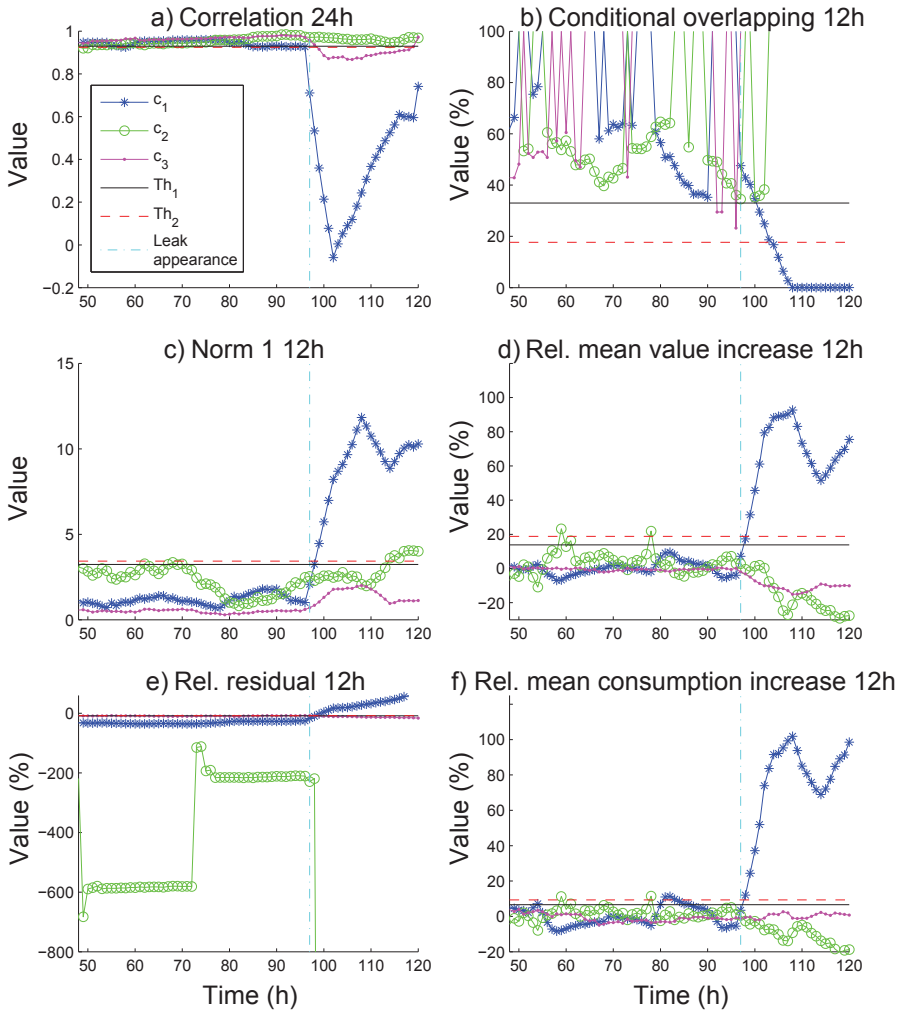

Figure 5.4: Detection indicators values and defined thresholds for the complete set of data (3 days for the training and 2 days for the evaluation). The leak starts at h=97

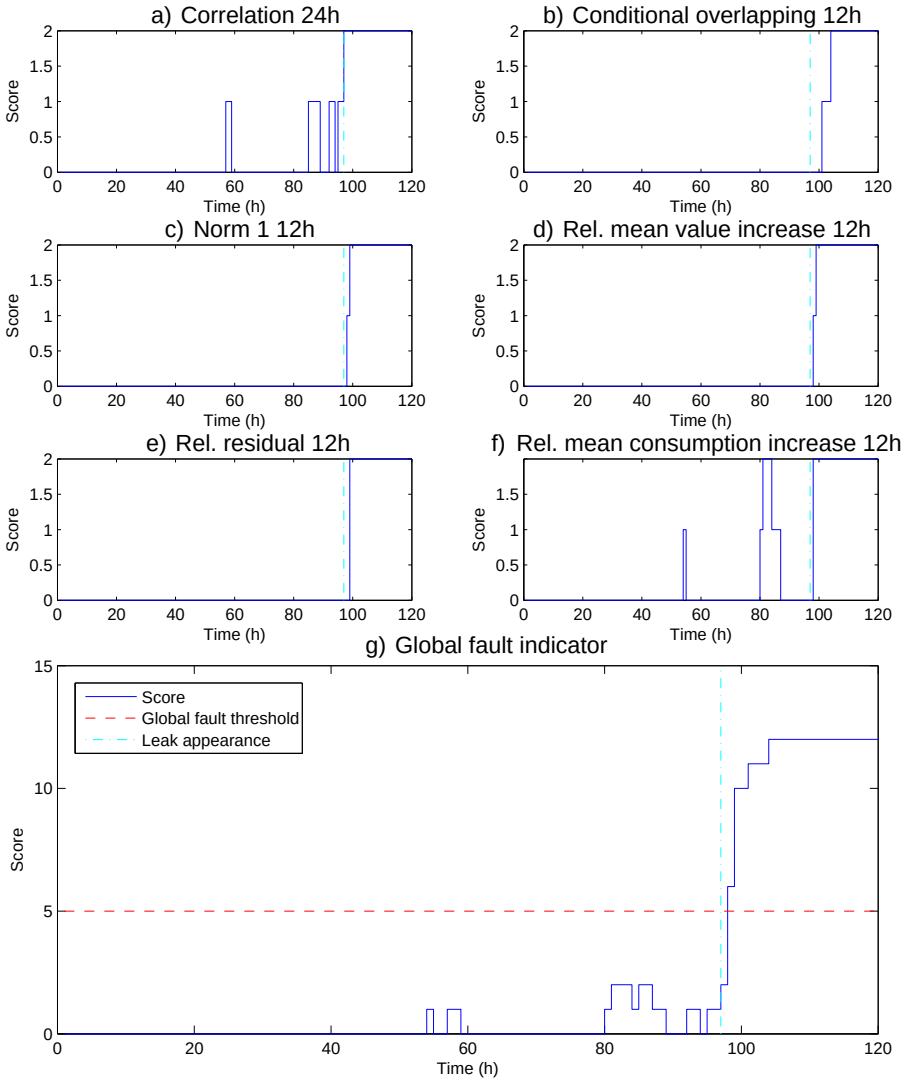

Figure 5.5: Evaluation of detection indicators for demand component $\mathbf{c}_1$

Table 5.1: Leak membership to each demand component computed from the demand components variations after the leak detection

$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$
79%	17%	4%

Table 5.2: Correlation between nodes memberships and the leak membership

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
-0.57	-0.35	0.77	-0.94	-0.98	0.37	**0.99**	-0.64	-0.64	-0.6

the leak starting at node '7', then node '3' and finally node '6'. The company would find the leak in the first indicated location.

The application of the *leak membership* method leads to the same result. Table 5.1 contains the leak membership that has been computed from the variations in the three demand components after the triggering of the detection alarm. The correlations between the leak memberships and each of the network nodes' memberships are presented in Table 5.2. It can be seen that node '7' has the highest correlation (0.99), followed by node '3' (0.77) and node '6' (0.37).

Chapter 6

Application to real scenarios

The collaboration in projects with the Barcelona water company AGBAR has permitted access to real data from three different networks: Canyars (2015), Nova Icària (2012), and Castelldefels Platja (2015). Results using these three networks are sorted depending on the network size.

First, calibration results using Canyars network are presented. The calibration is performed using all the available sensors, and evaluated with data from following weeks. Then, Nova Icària network is also calibrated, but leaving one of the sensors to perform the evaluation with the same scenario. Leak detection and localization results are presented using synthetic data, as real data were insufficient. Finally, Castelldefels Platja network is used as an example of the straightforward application of all the developed methodologies in any new WDN. The required engineering time is evaluated. Network reduction using skeletonization is performed in Nova Icària and Castelldefels Platja networks.

6.1 Canyars network

Canyars DMA is situated in Castelldefels (Catalonia) (Figure 6.1), supplying water to two neighborhoods: Castell-Poble Vell and Canyars. The network model is composed of 721 pipes and 698 junctions. Water is supplied from the transport network through a pressure reduction valve, depicted in Figure 6.2 with a blue triangle. Pressure and flow are monitored at the water inlet with a sample time of 10 min. The resolution is 0.3 l/s for the flow sensor, and 0.1 mwc (meters of water column) for the pressure sensor. The minimum night flow is of about 3 l/s, and the peak-hour flow is 27 l/s. Pressure control is applied to this network, fixing the pressure at the PRV outlet to 38 meters during night-time and 47

Figure 6.1: Canyars DMA location in the catalan city of Castelldefels. Satellite image extracted from Google Earth

meters during day-time. The average daily maximum head-loss between any two network junctions is 13.4 meters.

The organization of results in this section is as follows: Initially, the sampling design process is performed to choose the sensors to be installed, comparing the solution with the water utility final decision. Second, available data are analyzed and classified for the multiple calibration stages. Then, the parameterization process is performed considering the existence of the installed sensors. Finally, gross errors in sensors are corrected and the online calibration process applied. Results using multiple calibration horizons are compared to analyze the advantages and disadvantages of their use.

6.1.1 Sampling design

The parameterization and sampling design processes were performed to propose the location of three pressure sensors and the parameter definition for demand calibration, as explained in Chapter 3. However, the proposed sensors' locations differ from the final ones, which have been obtained from the methodology in Bonada et al. [2014] (based on leak detection), developed by Cetaqua, a water

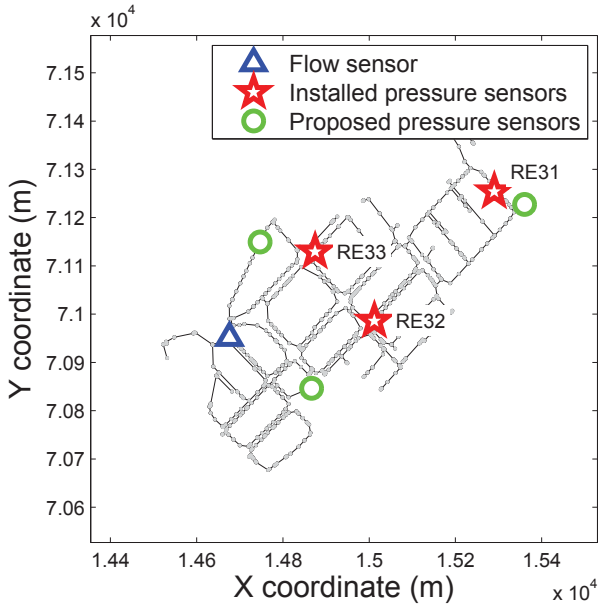

Figure 6.2: Canyars EPANET network model with highlighted sensors. The network water input is signaled with a blue triangle; the installed pressure sensors are signaled with red stars, and the proposed pressure sensors are signaled with green circles. The flow sensor is installed at the input pressure reduction valve, so that the total flow consumed in the network is known

Table 6.1: Canyars sensors sampling times (minutes), and offset corrections (meters)

Sensor ID	*RE33*	*RE31*	*RE32*
Sampling Time (min)	10	30	10
Offset correction (m)	0.19	0.05	0.4

technological center within the AGBAR group. The installed sensors can still be used to calibrate demands by defining the demand components depending on the available sensors' locations, thanks to the versatility of the proposed method (Section 3.3). Figure 6.2 depicts the proposed sensors locations with green circles, and the final locations with red stars. The resolution of the sensors is 0.1 mwc and the sampling times are defined in Table 6.1.

6.1.2 Data Analysis

Data from March 9th 2015 to March 13th 2015 (Monday-Friday) are used for the calibration process. Data from the following week, March 16th 2015 to March 20th 2015 (Monday-Friday), are used to validate and analyze the calibrated demand components. Previously, weekdays from March 3rd 2015 to March 6th

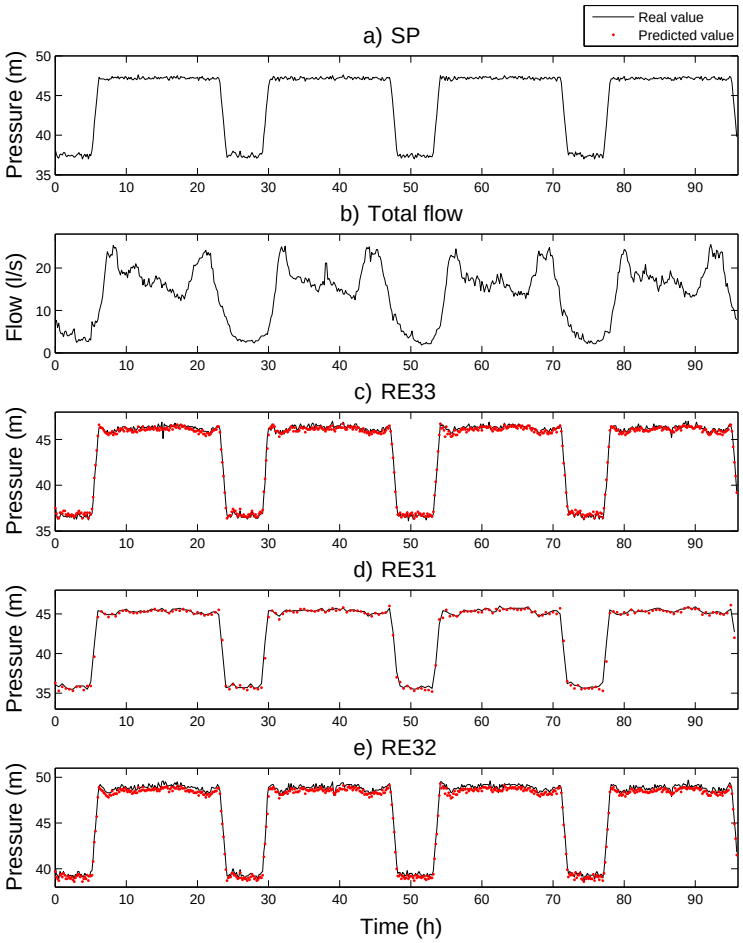

Figure 6.3: Canyars network real and predicted data from March 3rd 2015 to March 6th 2015 (precalibration week). Black lines and red dots refer to real and predicted data, respectively

2015 (Tuesday-Friday) are used to analyze and correct the data coming from the network, and to perform the parameterization process before the calibration starts. March 2nd 2015 (Monday) is not used due to missing data. These three weeks will be referred as precalibration week, calibration week and validation week. Weekends are not considered in this case study, but would follow the same calibration procedure as weekdays.

Figure 6.3 shows the complete set of data from the precalibration week, including boundary conditions (input valve's pressure set point and total flow) and the three pressure measurements. Black lines and red dots refer to real and predicted data, respectively. Predicted data have been obtained from simulating the network model with the given boundary conditions using the basic demand model presented in Equation 2.12. Figure 6.3.a shows the aforementioned pressure control at the DMA input.

Figure 6.4 shows the pressure prediction error in the three available sensors when using the basic demand model. The blue thin line corresponds to the raw error using all data, and the red thick line represents the smoothed error, which has been computed by means of a smoothing spline. The green dashed line corresponds to the mean pressure prediction error. This error is treated as an offset that cannot be associated to the demand model. As suggested in Pérez et al. [2014], the offset is corrected to eliminate possible depths errors, model nodes' elevations inaccuracies, roughness modeling errors, or badly calibrated sensors' offsets. The same correction in each sensor is also considered when using data from the calibration and validation weeks. Table 6.1 contains the specific correction for each sensor.

6.1.3 Parameterization

Data from the precalibration week are used to compute the sensitivity matrices to perform the parameterization process. The memberships of each nodal demand to three demand components are computed using Pseudo-code 1, considering the three installed sensors. Figure 6.5 depicts, in each of the network maps, the membership of each node to a particular demand component: the darker the node, the higher the membership to that component. Each map in Figure 6.5 also includes the location of the sensor with the highest sensitivity to the component drawn.

The average percentage of consumption $\overline{\mathbf{d}}_{\mathbf{c}_j}$ of demand component j is computed from the billing information (nodal base demands) and the recently computed memberships as:

$$\overline{\mathbf{d}}_{\mathbf{c}_j} = 100 \cdot \sum \mathbf{BDM} \cdot \mathbf{M}_{(:,j)} \tag{6.1}$$

Table 6.2 sums up the average percentage of consumption of each demand component: demand component $\mathbf{c}_2$ has the lowest percentage of consumption (18.6%), whereas $\mathbf{c}_1$ and $\mathbf{c}_3$ have both roughly a 40% of consumption. This information is used to analyze the calibration results: errors in the average percentage of consumption of the calibrated demand components compared to the assumed consumption in Table 6.2 can be assigned to background leakage, burst, fraudulent consumptions, unknown status valves, non-metered users, or wrong billing information.

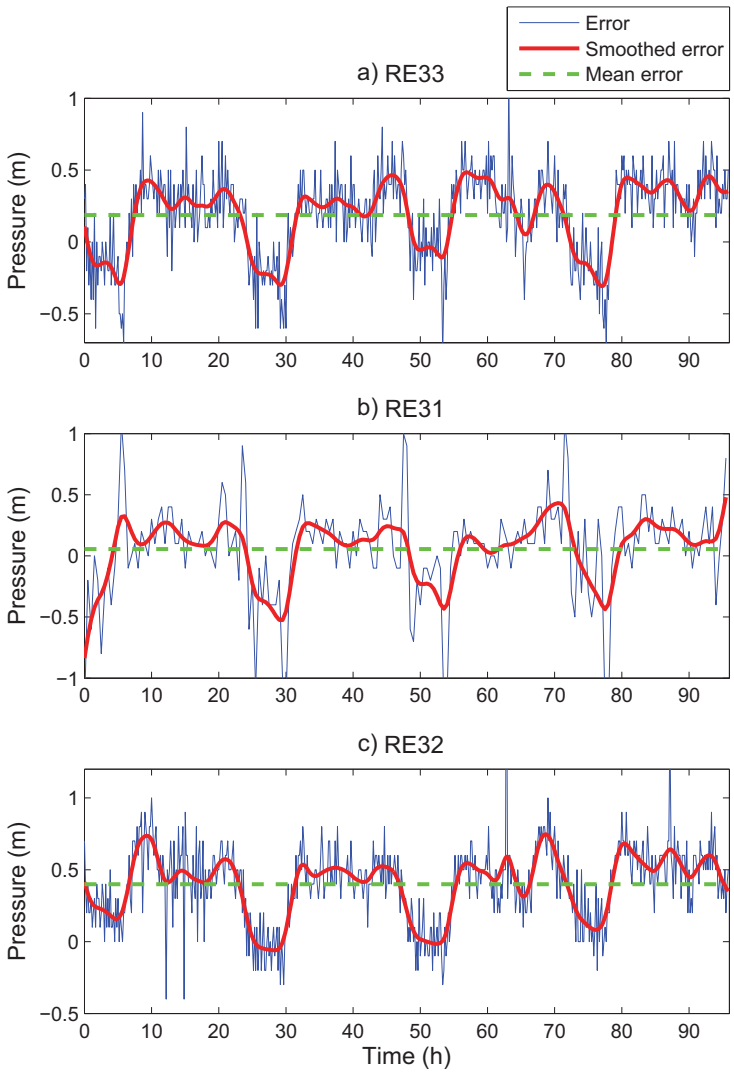

Figure 6.4: Canyars network pressure prediction error in the three installed sensors during the precalibration week using the basic demand model. The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

Table 6.2: Average percentage of demand components' water consumption in Canyars network computed from billing

Demand component	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$
Average percentage of water consumption	40,2%	18,6%	41,2%

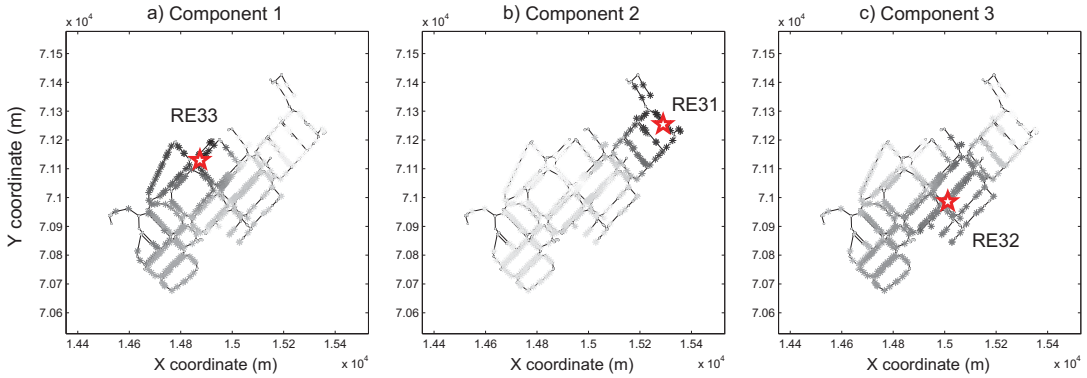

Figure 6.5: Memberships of nodes to each demand component in Canyars network considering the three available sensors. Each representation of the network depicts a grayscale map with the membership of each node to a particular demand component: the darker the node in the map, the higher the membership of the node to the demand component. The sensor with the highest sensitivity to variations in each demand component is also depicted in each map

6.1.4 Calibration Results

Once the parameters that are going to be calibrated are defined, and the information from the network sensors has been thoroughly analyzed and corrected, the online calibration starts. In the current study, the online part is not actually online, though the process and the execution are exactly the same.

Three different tests are considered. In each test, demand components values are calibrated with data from the calibration week, and validated with data from the validation week. Furthermore, a new set of demand components values is calibrated using data from the validation week to compare the difference between updating the demand components values or keeping the same values once they are calibrated. Each test considers a different value of the calibration horizon parameter H_c, which defines, for each sensor, the number of days from which data for a specific hour are going to be used to calibrate the demand components values for that particular hour:

1. Test 1: $H_c = 5$ days. The online calibration starts on March 13th at 0 am (Friday). At each hour, data from the current day and four previous days (Monday-Thursday) are used to calibrate a unique value for each demand component. The online calibration continues until March 13th at 23 pm (same day), when a total of 24 values will have been calibrated for each demand component. The calibrated demand components values at a specific hour minimize the mean prediction error from that hour of the five days

used to calibrate them.

2. Test 2: $H_c = 1$ day. The online calibration starts on March 9th at 0 am (Monday). At each hour, only data from the current day are used to calibrate the demand components values. The online calibration continues until March 13th at 23 pm, when a total of 120 values will have been calibrated for each demand component. The calibrated demand components values at a specific hour minimize exclusively the prediction error for the hour of the day used to calibrate them.

3. Test 3: $H_c = 3$ days. The online calibration starts on March 11th at 0 am (Wednesday). At each hour, data from the current day and two previous days are used to calibrate a single value for each demand component. The calibration continues until March 13th at 23 pm, when a total of 76 values will have been calibrated for each demand component. The calibrated demand components values at a specific hour minimize the mean prediction error from that hour of the three days used to calibrate them. The first 24 demand components values minimize the error in data from Monday to Wednesday, demand components values from 25h to 48h minimize the error in data from Tuesday to Thursday, and demand components values from 49h to 72h minimize the error in data from Wednesday to Friday.

Figure 6.6 presents an example of the calibrated demand component $\mathbf{c}_2$ daily values during the calibration week: a) demand component values calibrated with $H_c = 5$, that minimize the whole week pressure prediction error; b) demand component values calibrated with $H_c = 1$, that minimize the pressure prediction error during Friday; and c) demand component values calibrated with $H_c = 3$, that minimize the pressure prediction error for three consecutive days. Demand component values are presented as white circles surrounded by the 95% confidence intervals (green boundaries).

Table 6.3 collects the pressure prediction root mean squared error (RMSE) for each test. Each row corresponds to a different test. The three main columns represent the data used, and the use itself: column 1 contains the calibration results for the calibration week, column 2 contains the validation of these results using data from the validation week, and column 3 contains the results considering a new calibration during the validation week. Each main column is composed of three subcolumns: the first subcolumn contains the pressure prediction RMSE when using the basic demand model, the second subcolumn contains the pressure prediction RMSE when using the calibrated demand components model, and

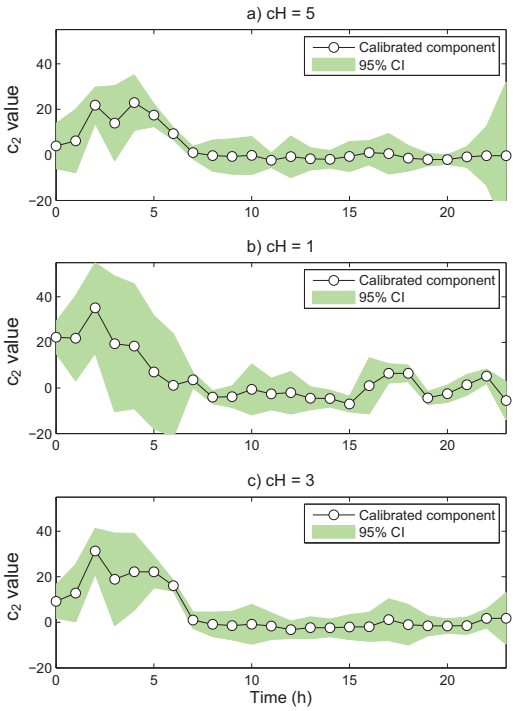

Figure 6.6: Canyars calibrated demand component $\mathbf{c_2}$ with 95% confidence intervals (CI): a) $H_c = 5$, b) $H_c = 1$ during 24 hours, and c) $H_c = 3$ during 24 hours

the third subcolumn contains the error improvement when using the demand components model instead of the basic model, expressed as a relative percentage.

Figure 6.7 presents a comparison between the pressure prediction error in the calibration week using the basic demand model and the calibrated demand components model, both from Test 2 ($H_c = 1$). The depths corrections in Table 6.1 have been used to correct the sensors' offsets. The columns of subfigures correspond to the basic demand model and demand components model, respectively. Each row of subfigures corresponds to each of the three sensors. The blue lines correspond to the raw errors using all data, and the red lines represent the smoothed errors, which have been computed by means of a smoothing spline. The green dashed lines correspond to the mean pressure prediction errors.

Finally, Figure 6.8 presents the average percentage of consumption computed for each demand component. The black line corresponds to the assumed percentage previously shown in Table 6.2, and the red lines correspond to the demand components' average percentage of consumption in the nine scenarios presented in Table 6.3.

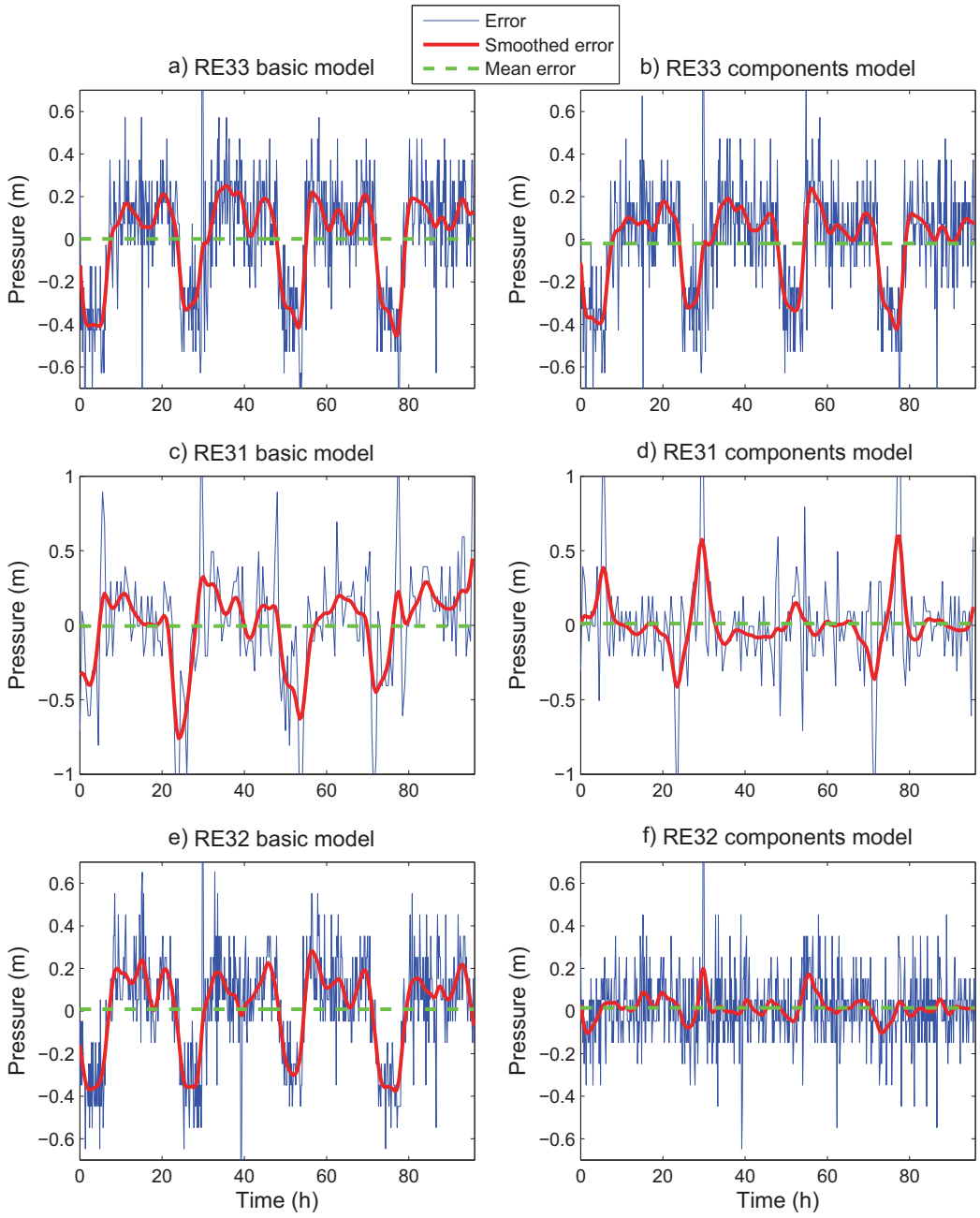

Figure 6.7: Canyars network pressure prediction error during the calibration week using the basic demand model (column 1) and the demand components model (column 2) using $H_c = 1$. The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error. The depth correction has been applied to all sensors

Table 6.3: Canyars pressure prediction RMSE and percentage of improvement in the three calibration tests performed. Rows indicate the calibration horizon used, while main columns correspond to the data used and the objective of the analysis: the first column presents the calibration results using data from the calibration week, the second column presents the validation of the first column results for the validation week, and the third column presents the calibration results for the validation week. Within each main column, first and second subcolumns refer to the pressure prediction RMSE using the basic and demand components model, respectively; and third subcolumn presents the error improvement obtained when using the demand components model instead of the basic model

	Calibration RMSE (m) March 9th - 13th			**Validation RMSE (m) March 16th - 20th**			**Calibration RMSE (m) March 16th - 20th**		
Test	**Basic**	**Comp.**	**Imprv.**	**Basic**	**Comp.**	**Imprv.**	**Basic**	**Comp.**	**Imprv.**
$H_c = 5$	0,298	0,267	11%	0,304	0,282	7%	0,304	0,277	9%
$H_c = 1$	0,298	0,250	16%	0,304	0,296	2%	0,304	0,252	17%
$H_c = 3$	0,294	0,254	13%	0,296	0,294	1%	0,296	0,267	10%

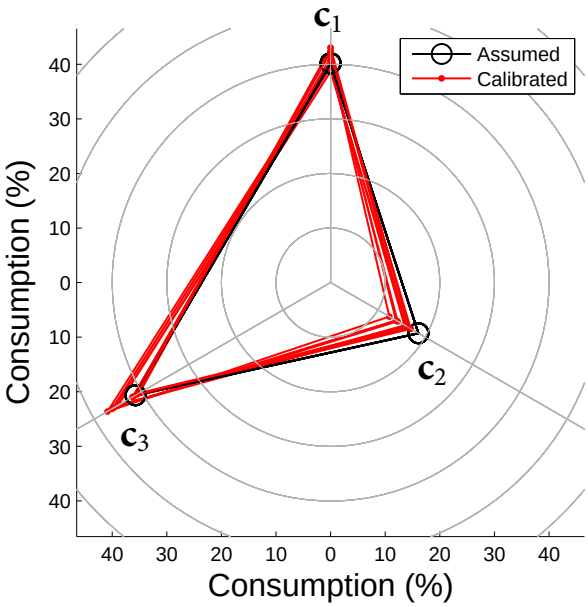

Figure 6.8: Demand components' average percentage of consumption assumed from billing (black line with circles), and obtained with the calibrated demand components in the nine scenarios presented (red lines with dots)

6.1.5 Discussion

The basic demand model has been used to simulate the network during the pre-calibration week. Figure 6.4 shows that the pressure prediction error when using the basic demand model follows the profile of the daily total consumption in Figure 6.3.b, as this demand model is not able to assign a different behavior to each zone of the network. The prediction error can be improved by using of the demand components model, which allows having multiple demand behaviors depending on the location of nodes in the network.

Results in Table 6.3 show that the use of calibrated demand components to model nodal demands minimizes the RMSE in the predicted pressures in the three tests performed, compared to the basic demand model. This is verified through the positive prediction error improvement in all tests. The comparison of the pressure prediction error between the basic demand model and the demand components model is shown in Figure 6.7. The pressure prediction error is reduced in the three sensors. However, the prediction improvement in each sensor is different. The proximity of sensor 'RE33' to the input, where the PRV sets the pressure level, can explain the little improvement. On the other hand, the pressure prediction error in sensors 'RE31' and 'RE32' is minimized considerably. Peaks in the smoothed prediction error in Figure 6.7.d are due to single large errors in pressure predictions due to the PRV set point abrupt changes from night to day and vice versa. Summing up, the calibrated demand components model smoothens the total water consumption-like profile of the pressure prediction error, as each area of the network has now a distinct behavior.

The minimization of the prediction error depends on the calibration horizon used: the lower H_c is, the higher the improvement on the predicted pressures. In both first and third main columns, Test 2 ($H_c = 1$) has the greater improvement. Using data exclusively from one day allows the demand components to focus on minimizing a single error. If demand behaviors change in the real network, Test 2, which uses $H_c = 1$, captures instantaneously these changes. On the other hand, Tests 1 and 3 accommodate the demand components values more slowly, since at each sample there are multiple prediction errors to be minimized at the same time. This is also observed in Figure 6.6, where the profile from Test 2 ($H_c = 1$) captures the daytime demand variations with more detail, opposite from Test 1 and 3, where the day-time profile is flatter.

Nevertheless, the validity of the demand components calibrated with higher H_c values is longer, allowing the demand components to be applied in future

data and still minimize the prediction error. That is because the use of multiple measurements to calibrate a unique set of demand components values leads to demand components that represent the usual behavior in the network, ignoring special events in particular days that are not repetitive from one week to another. This fact is checked in the second main column in Table 6.3, where Test 1, that uses $H_c = 5$, is the one with less error when using the demand components calibrated with data from the previous week.

The comparison between the second and third main columns in Table 6.3 shows that continuously updating the demand model with new online calibrated demand components is better than calibrating a set of demand components and using them with new data, in terms of pressure prediction RMSE. For example, in Test 2, the pressure prediction RMSE during the validation week (second main column) improved 2% compared to the basic model; on the other hand, if the online calibration is performed, the improvement on the pressure prediction RMSE is 17% (Test 2 in third main column). Hence, the continuous online calibration is required to capture the daily changes in the network demand behaviors, such as changes in users habits or new users coming online. Besides, the capture of changes in the network demand behaviors provoked by leaks, unnoticed changes in valves status, or other special events, can be analyzed and used to detect and locate them, as explained in Chapter 5 and applied in subsection 6.2.3.

The uncertainty of the calibrated demand components can be analyzed through Figure 6.6. During night-time, due to the nonlinear headloss/flow relation (see Equation 2.2), pressure is less sensitive to demand variations. Consequently, the singular values of the sensitivity matrix are smaller, increasing the calibrated parameters variance as seen in Equation 4.7. Using extra data to calibrate demand components (Figure 6.6.a and .c) reduces the uncertainty in the calibrated demand components values compared to using data only from the current day (Figure 6.6.b).

Finally, Figure 6.8 shows that the average percentage of consumption of each demand component in the nine scenarios tested is maintained. The variations observed (5%-6%) respecting to the assumed percentage of consumption can be associated with the missing information from weekend demand components, which have not been calibrated in this work, small background leakage or fraudulent consumptions.

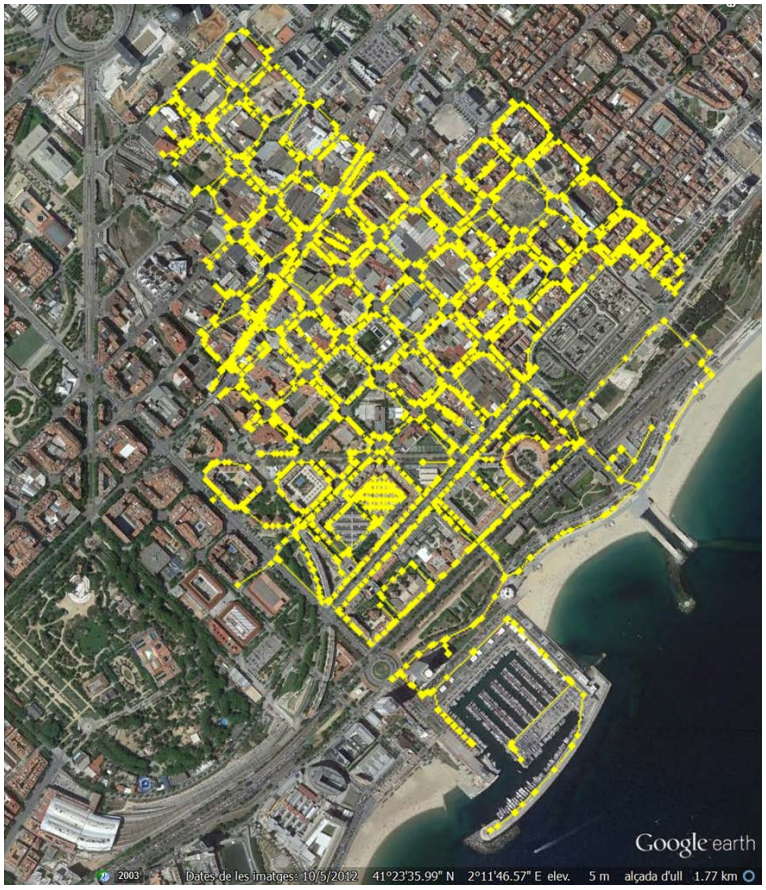

Figure 6.9: Nova Icària DMA location in the catalan city of Barcelona. Satellite image extracted from Google Earth

6.2 Nova Icària network

Nova Icària DMA is situated in Barcelona (Catalonia) (Figure 6.9), supplying water to three neighborhoods: la Vila Olímpica del Poblenou, el Poblenou and El Parc i la Llacuna del Poblenou. It is composed of 3455 pipes and 3377 junctions. Water is supplied from the transport network through two pressure reduction valves, presented in Figure 6.11 with blue triangles. Pressure and flow are monitored at both water inlets with a sample time of 10 min. The resolution is 0.3 l/s for the flow sensors, and 0.1 mwc for the pressure sensors. The minimum night flow is of about 20 l/s, and the peak-hour flow is 70 l/s. The pressure fixed at the input PRVs are of about 50 meters. The average daily maximum head-loss between any two network junctions is 6.8 meters.

The high number of junctions and pipes in the network model increases considerably the computational effort required when applying any methodology using

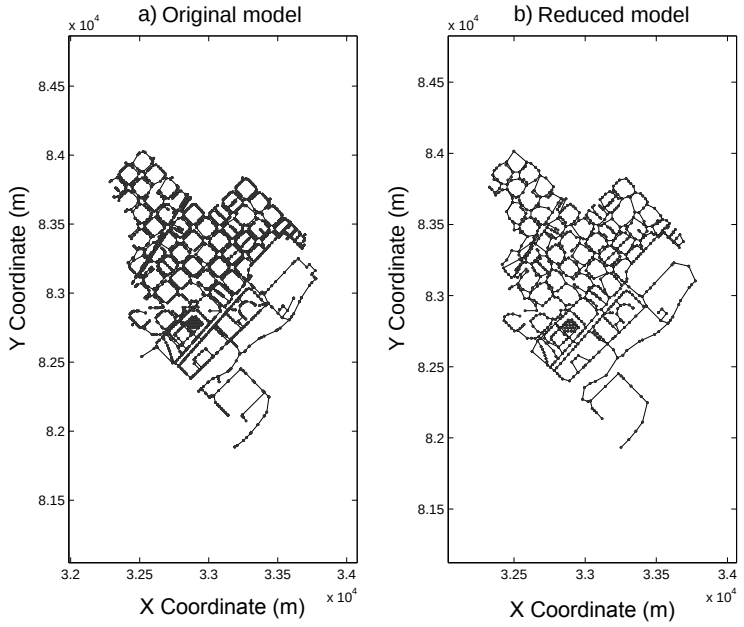

Figure 6.10: Nova Icària network original and reduced EPANET models

it. For this reason, the skeletonization process presented in Appendix A is applied to the network model, considering only those reductions that do not affect the hydraulic behavior of the network. Figure 6.10 depicts the original model and the reduced model. The number of junctions has been reduced from 3377 to 1520 (45%), and the number of pipes from 3455 to 1644 (47.5%). The reduced model has been evaluated by comparing the pressures in the 1520 remaining nodes with the same nodes in the original network model. The highest pressure error during 144 samples (one day, with a sampling time of 10 minutes) is 0.000003 meters, significantly lower than the sensors' resolutions (0.1 m).

6.2.1 Context of the work

Three different works have been done using this network:

a) Demand component calibration using real data, presented in subsection 6.2.2.

b) Leak detection and localization through demand component calibration using synthetic data, presented in subsection 6.2.3.

c) Leak detection and localization through demand component calibration using real data, presented in subsection 6.2.4.

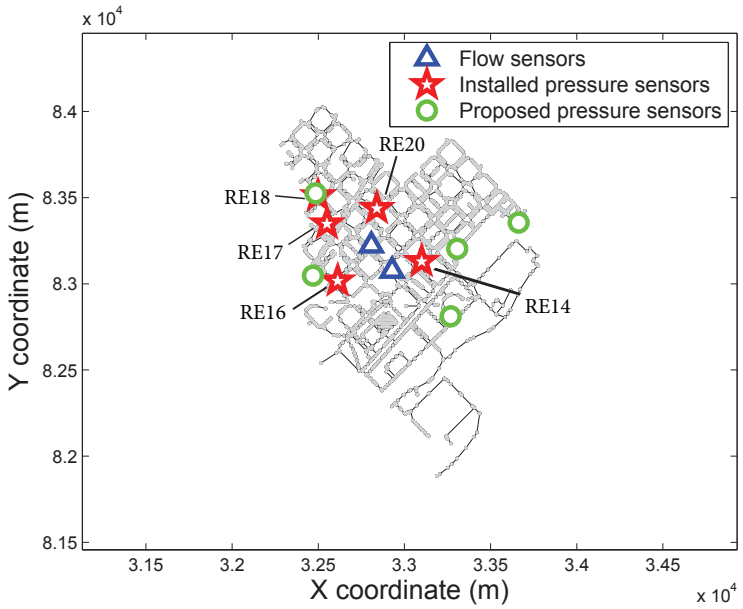

Figure 6.11: Nova Icària EPANET network model with highlighted sensors. The network water inputs are signaled with a blue triangle; the installed pressure sensors are signaled with red stars and the proposed pressure sensors are signaled with green circles. The flow sensors are installed at the input pressure reduction valves, so that the total flow consumed in the network is known

Table 6.4: Nova Icària sensors sampling times (minutes), and offset corrections (meters)

Sensor ID	*RE16*	*RE20*	*RE17*	*RE18*	*RE14*
Sampling Time (min)	10	10	10	10	10
Offset correction (m)	2.17	1.01	1.35	2.19	1.32

6.2.2 Demand calibration with real data

During the RTNM project, Nova Icària was the most used DMA to test the methodologies developed. The availability of this network has been useful to test the demand calibration methodology in multiple scenarios. Five pressure sensors were installed in this network for leak detection and localization, using the methodology presented in Pérez et al. [2009]. The location of these sensors differ from the one obtained by the methodology proposed in Chapter 3. Figure 6.11 depicts the desired sensors locations with green circles, and the real locations with red stars. The resolution of the installed sensors is 0.1 mwc and the sampling times are defined in Table 6.4.

Data analysis

Data from December 10th 2012 (Monday) to December 14th 2012 (Friday) are used to calibrate the network model. Previously, days from November 21st 2012 (Wednesday) to November 25th 2012 (Sunday) are used to analyze and correct the data coming from the network, and to perform the parameterization process before the calibration starts. These two sets of data will be referred to as precalibration data and calibration data.

Figure 6.12 shows the pressure prediction error in the five available sensors when using the basic demand model for the precalibration dataset. The blue line corresponds to the raw error using the filtered data, and the red line represents the smoothed error, which has been computed by means of a smoothing spline. The green dashed line corresponds to the mean pressure prediction error. This error is treated as an offset that cannot be associated to the demand model. As suggested in Pérez et al. [2014], the offset is corrected to eliminate possible depths errors, model nodes' elevations inaccuracies, roughness modeling errors, or badly calibrated sensors' offsets. The same correction in each sensor is also considered when using data from the calibration dataset. Table 6.4 contains the specific correction for each sensor.

Parameterization

For this test, only four parameters will be considered, related with four sensors (RE14, RE16, RE17, and RE18). The fifth sensor (RE20) will be used to validate the calibrated demand model.

Data from the precalibration set are used to compute the sensitivity matrices to perform the parameterization process. The memberships of each nodal demand to four demand components are computed considering the four installed sensors, as described in Pseudo-code 1. Figure 6.13 depicts, in each of the network maps, the membership of each node to a particular demand component: the darker the node, the higher the membership to that component. Each map in Figure 6.13 also includes the location of the sensor with the highest sensitivity to the component drawn.

Table 6.5 sums up the average percentage of consumption (Equation 6.1) of each demand component: demand component $\mathbf{c}_1$ and $\mathbf{c}_4$ have the lowest percentages of consumption (14.6% and 16,1%, respectively), as they cover small areas in the DMA. On the other hand, demand components $\mathbf{c}_2$ and $\mathbf{c}_3$ have a 38% and 31% of consumption, respectively, as they cover larger areas. This informa-

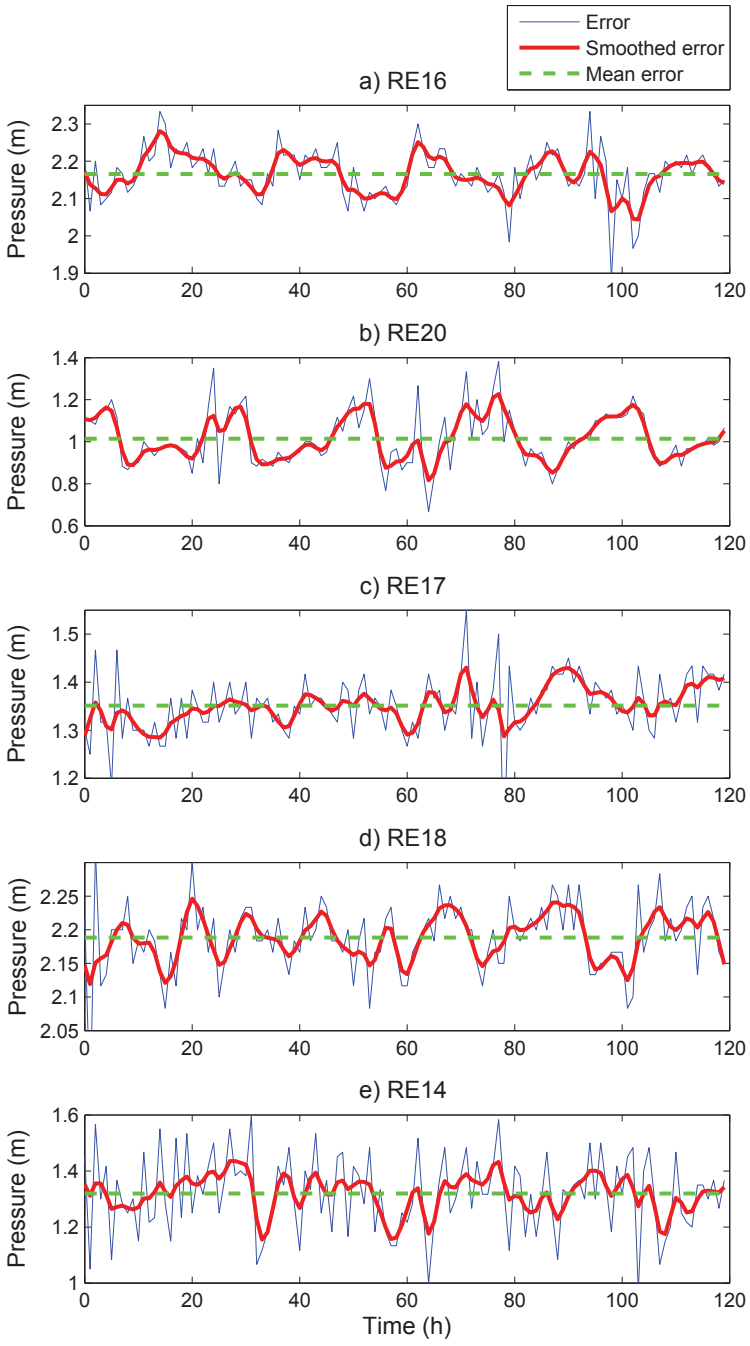

Figure 6.12: Nova Icària network pressure prediction error in the five installed sensors. The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

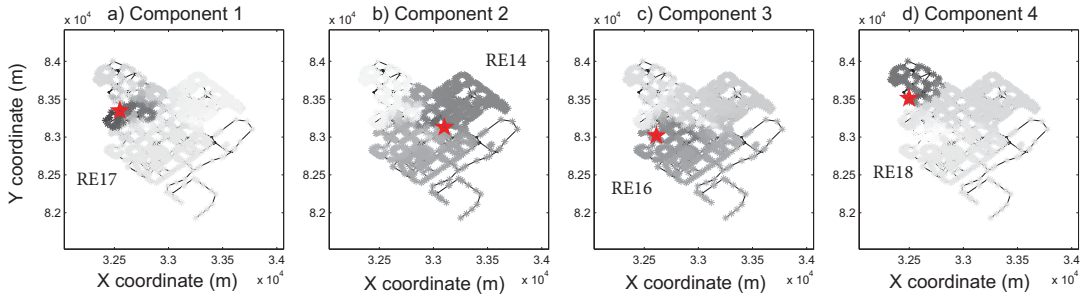

Figure 6.13: Memberships of nodes to each demand component in Nova Icària network considering the four selected sensors. Each representation of the network depicts a grayscale map with the membership of each node to a particular demand component: the darker the node in the map, the higher the membership of the node to the demand component. The sensor with the highest sensitivity to variations in each demand component is also depicted in each map

Table 6.5: Average percentage of demand components' water consumption in Nova Icària network computed from billing

Demand component	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\mathbf{c}_4$
Average percentage of water consumption	14,6%	38%	31,3%	16,1%

tion is used to analyze the calibration results: errors in the average percentage of consumption of the calibrated demand components compared to the assumed consumption in Table 6.5 can be assigned to background leakage, burst, fraudulent consumptions, unknown status valves, non-metered users, or wrong billing information.

Genetic algorithms settings

The results presented in the following section, which have been obtained with the least square procedure, will be compared with the results obtained by means of an evolutionary optimization method (Section 2.3): Genetic Algorithm (GA). Matlab includes a Genetic Algorithm toolbox, which has been used for the current test. The use of GA involves the setting of a series of parameters that are used during the calibration process. The following list presents the settings for the GA:

- Cost function: The cost function in Equation 4.17 has been used.
- Generations: 10.
- Population size: 20.

Table 6.6: Nova Icària pressure and flow prediction RMSE using the basic demand model and the demand components model calibrated with LS and GA. For each of the calibration methods, the percentage of improvement of the demand components model with respect to the basic demand model is also presented. The pressure RMSE has been calculated for the four sensors used during the calibration process

	Basic Model	**Demand Components Model**			
		Least Squares		**Genetic Algorithm**	
	RMSE	**RMSE**	**Improvement**	**RMSE**	**Improvement**
Pressure (m)	0.387	0.385	0.5%	0.407	-5.1%
Flow (l/s)	6.073	5.416	10.8%	3.615	40%

- For each hour, the genetic algorithm is performed five times, and the best result out of the five results is chosen.

Calibration Results

The calibration process presented in Chapter 4 has been applied to the presented network, considering the four mentioned pressure sensors and the defined demand components. A calibration horizon $H_c = 5$ has been used, i.e. data from the same hour of five different days have been used to calibrate the demand component values for that specific hour, minimizing the prediction error at the five days simultaneously. Samples within the same hour have been previously filtered to reduce noise. Two different optimization methods have been used to perform the same calibration: Least squares minimization using the SVD to compute the inverse of the sensitivity matrix, and GA.

Figure 6.14 presents the calibrated demand components daily values, which are valid for each of the five days in the calibration dataset. Subfigures a)-d) correspond to the LS solution, whereas subfigures e)-h) correspond to the GAs solution. Demand component values are presented as white circles surrounded by the 95% confidence intervals (green boundaries).

Table 6.6 collects the pressure prediction RMSE (first row), and the flow prediction RMSE (second row), for each optimization method used. The percentage of error improvement has also been computed for each method and type of measurement, which represents the error improvement when using the demand components model instead of the basic demand model. Table 6.7 presents the same results for the fifth sensor, which has been used to evaluate the effect of the calibration on a non-used sensor.

Table 6.8 presents the value of the cost function (Equation 4.17) for each of

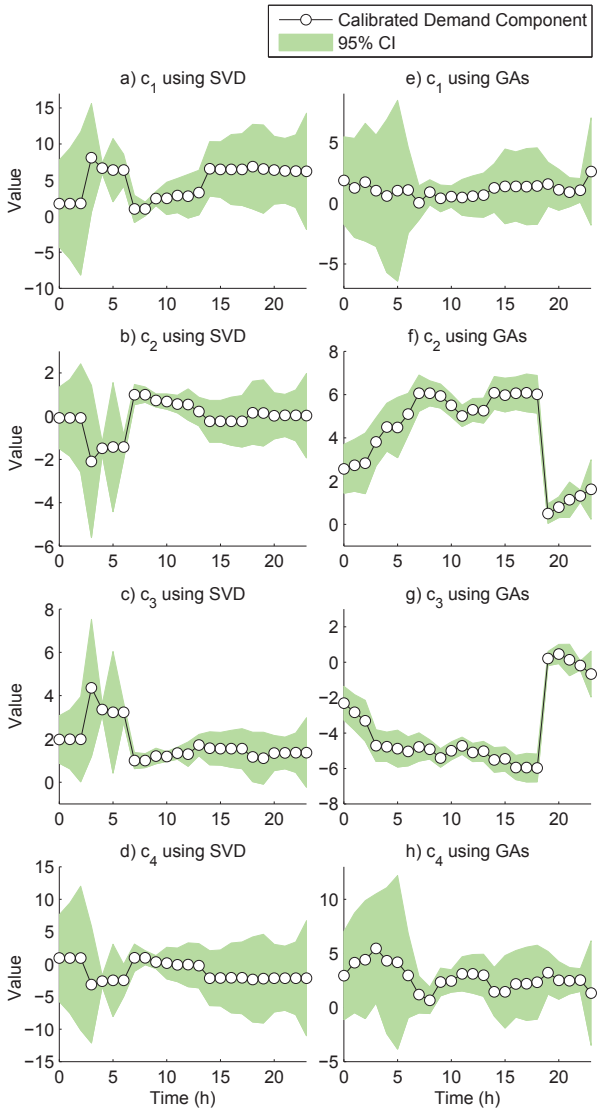

Figure 6.14: Nova Icària calibrated demand components with 95% confidence intervals (CI) using the LS-optimization [a), b), c), and d)], and GAs [e), f), g), and h)]

Table 6.7: Nova Icària pressure and flow prediction RMSE using the basic demand model and the demand components model calibrated with LS and GA. For each of the calibration methods, the percentage of improvement of the demand components model with respect to the basic demand model is also presented. The pressure RMSE has been calculated with the fifth sensor to evaluate the effect of the calibration on non-used sensors

	Basic Model	**Demand Components Model**			
		Least Squares		**Genetic Algorithm**	
	RMSE	**RMSE**	**Improvement**	**RMSE**	**Improvement**
Pressure (m)	0.282	0.281	0.1%	0.414	-46.8 %

Table 6.8: Nova Icària Equation 4.17 cost function evaluation using the basic demand model and the demand components model calibrated with LS and GA. For each of the calibration methods, the percentage of improvement of the demand components model with respect to the basic demand model is also presented

Basic	**Demand Components Model**			
Model	**Least Squares**		**Genetic Algorithm**	
$\mathcal{J}$	$\mathcal{J}$	**Improvement**	$\mathcal{J}$	**Improvement**
11.922	7.471	37.3%	0.686	94.24%

the calibration methods considered.

Figure 6.15 and Figure 6.16 depict the flow and pressure prediction errors, respectively, for each of the network sensors. Each figure has three columns of subfigures, corresponding to the prediction error when using the basic demand model (1st column of subfigures), the prediction error when using the demand components model calibrated using the LS method (2nd column of subfigures), and the prediction error when using the demand components model calibrated using the GAs (3rd column of subfigures).

Finally, Figure 6.17 presents the average percentage of consumption of each demand component, depending on the model and method used: the black line corresponds to the basic demand model (assumed demands from billing), the green line with triangles corresponds to the demand components model calibrated using the LS method, and the red line with circles corresponds to the demand components model calibrated using the GAs.

Discussion

The basic demand model has been used to simulate the network with the precalibration dataset. Figure 6.12 shows that the pressure prediction error when using the basic demand model follows a demand-like pattern, as this demand model is not able to assign a different behavior to each zone of the network. The prediction error can be improved by the use of the demand components model, which allows having multiple demand behaviors depending on the location of nodes in the network.

Results in Table 6.6 show that the use of demand components to model nodal demands minimizes the RMSE in the predicted flows considerably, specially when using the GAs (40% improvement). The pressure prediction RMSE is similar to the one obtained with the basic demand model when using the LS method. For the GA-based calibration, the pressure RMSE increases by 5% compared to the basic model. This pressure error increases when using the GAs is provoked by

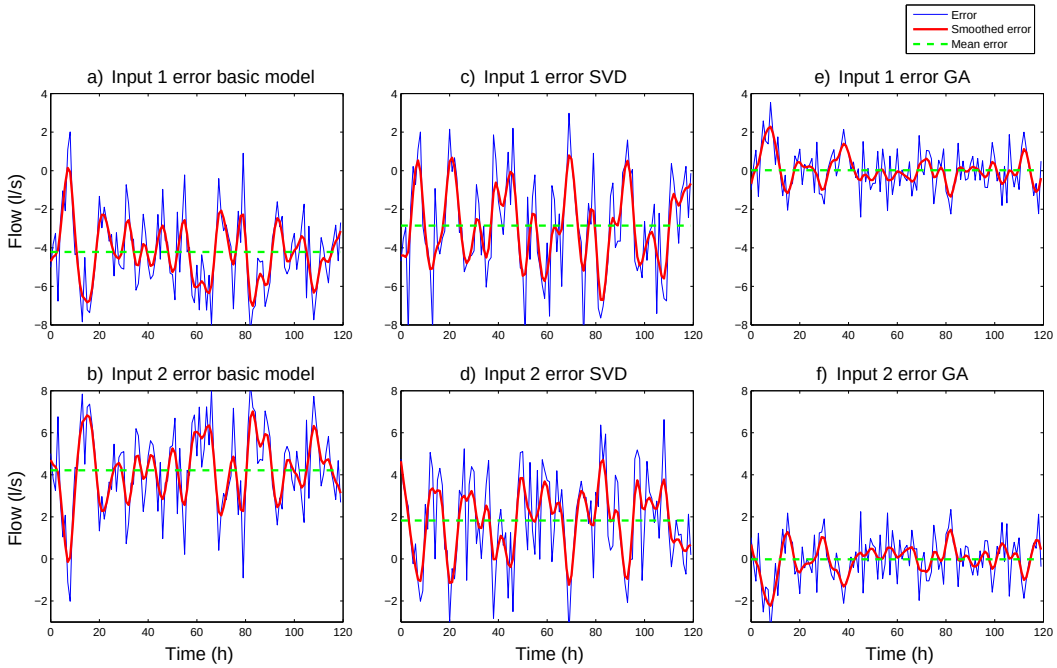

Figure 6.15: Nova Icària network flow prediction error at each network input (rows of subfigures) using the basic demand model [a) and b)], the demand components model calibrated using LS [c) and d)], and the demand components model calibrated using GAs [e) and f)]. The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

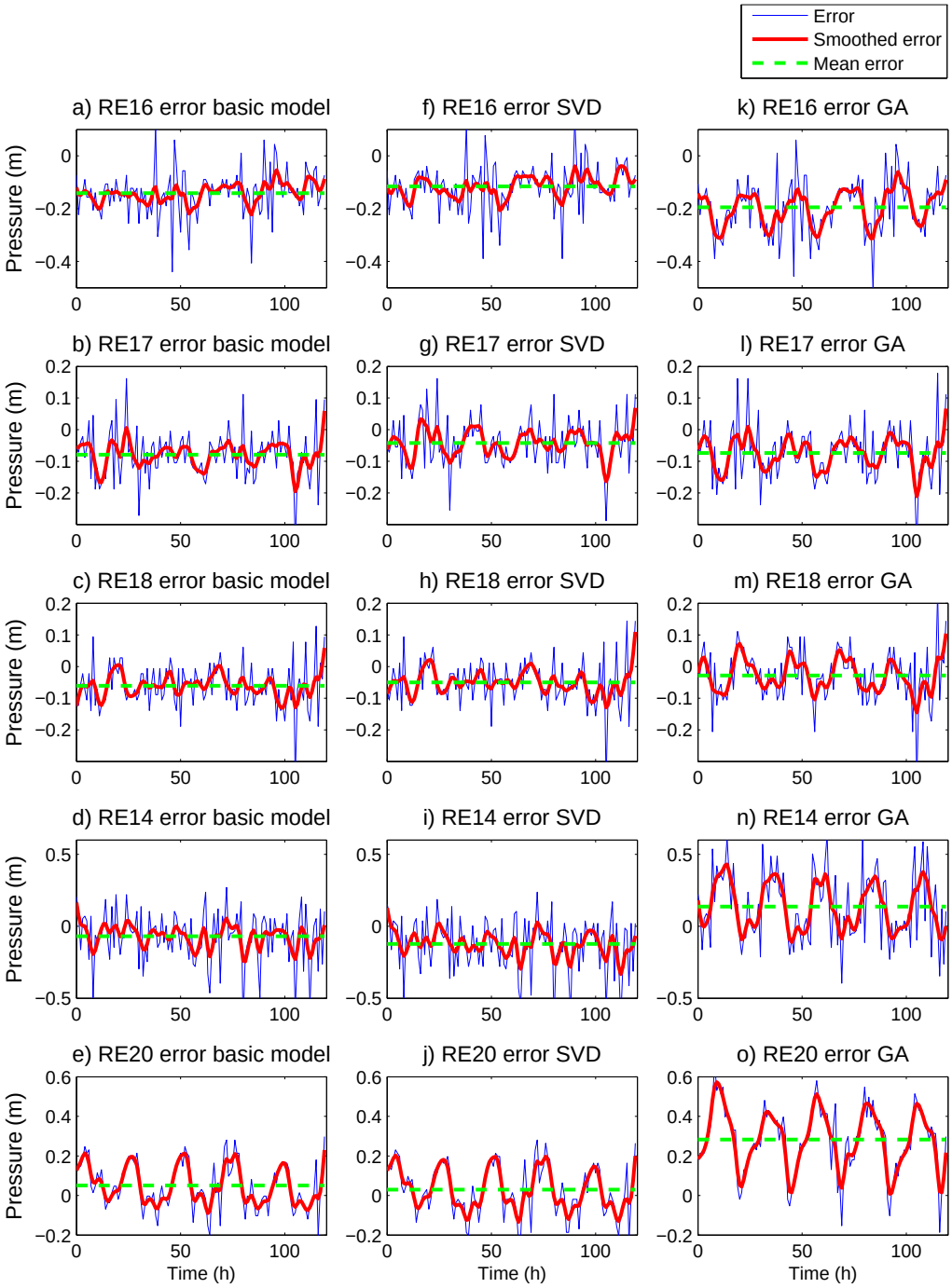

Figure 6.16: Nova Icària network pressure prediction error at each sensor (rows of subfigures) using the basic demand model (first column of subfigures), the demand components model calibrated using LS (second column of subfigures), and the demand components model calibrated using GAs (third column of subfigures). The last row of subfigures corresponds to the evaluation sensor, which has not been used during the calibration process. The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

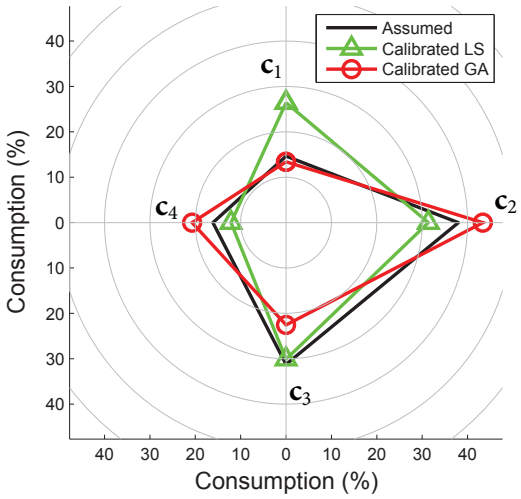

Figure 6.17: Average percentage of demand components' water consumption computed from billing (black line), LS-based calibration (green line with triangles), and GAs-based calibration (red line with circles) in Nova Icària network

the great improvement in the flow error, which minimizes the cost function even if the pressure error increases a little, as seen in Table 6.8.

The evaluation of results with the fifth sensor (RE20) presented in Table 6.7 show a tiny improvement in the pressure prediction error of the evaluation sensor when using the LS method, which means that the hydraulic behavior is consistent. On the other hand, the predicted pressure error in the evaluation sensor for the calibrated demand components using the GA increases by 46.8% compared to the basic demand model. This increase in the pressure error of the evaluation sensor is logical taking into account that even pressure measurements included in the cost function had increased.

Figure 6.15 and Figure 6.16 show a graphical representation of the previous results. It can be seen that the calibrated demand components using LS improve the flow error while keeping a similar pressure error. The GA generates a demand components model that reduces drastically the error in the predicted flows at the expense of pressure predictions.

The uncertainty of the calibrated demand components can be analyzed through Figure 6.14. During night-time, due to the nonlinear headloss/flow relation (see Equation 2.2), pressure is less sensitive to demand variations. Consequently, the singular values of the sensitivity matrix are smaller, increasing the calibrated parameters variance as seen in Equation 4.7. Besides, demand components with higher percentage of consumption (Table 6.5) have less uncertainty than demand components with low percentage of consumption. No substantial differences are

observed between the LS and the GA results in terms of uncertainty in the calibrated demand components.

Finally, Figure 6.17 shows that the LS calibration maintains the average percentage of consumption of demand component $\mathbf{c}_3$, which consumes 1/3 of the network water. The consumptions of $\mathbf{c}_2$ (the demand component with highest consumption) and $\mathbf{c}_4$ are decreased in favor of $\mathbf{c}_1$. On the other hand, the GA maintains the average percentage of consumption of demand component $\mathbf{c}_1$, while decreasing the consumption of $\mathbf{c}_3$ in favor of $\mathbf{c}_2$ and $\mathbf{c}_4$. The average percentage inaccuracies could be provoked by the inclusion of weekend data in the assumed values, whereas the calibrated data have been obtained only from weekdays. Note that Nova Icària is a touristic zone near the Barcelona port, with plenty of hotels, night clubs and restaurants, which change significantly their water consumption from weekdays to weekends.

In conclusion, the calibration using the LS method (as in Section 6.1) significantly improves the prediction error in the distribution of flows at the inputs, while keeping a similar pressure prediction as the basic demand model. Considering these results, the average percentage of consumption obtained with this calibration method would signal a bad spatial distribution of demands, which can indicate the presence of a leak (in this case, near demand component $\mathbf{c}_1$), or wrong information in billing data (missing consumers, fraudulent users, etc.).

The results obtained with the GA show that the flow error has been too much favored by the cost function, as it is checked in Table 6.8. A redefinition of the cost function could lead to better results, as the GAs give a better solution than LS in terms of cost function minimization due to the non-convexity of the optimization problem. However, the author suggests using the LS method which adopts the SVD for the resolution of the inverse problem for the proved good performance together with the faster resolution of the problem than the GA, which cannot be performed in a real time context.

6.2.3 Leak detection and localization with synthetic data

The second test presented in this section consists in the application of the leak detection and localization methodology presented in Chapter 5 to the Nova Icària network model using synthetic data. This test with synthetic data has been performed to check the validity of the fault detection method in a large network before its application into a real scenario.

Reality generation

The generation of synthetic data requires a previous emulation of reality. A complete set of synthetic demands has been computed to represent reality, where different consumers use water differently (e.g. household, commercial, industrial, etc.). First, ten diurnal demand patterns have been defined, representing different types of users. Each nodal demand in the network has an associated type of user. These types are mixed all over the network, emulating the real behavior of the used DMA. All patterns, and consequently all nodal demands, have different behaviors during weekdays and weekends. A random normal noise $N(0, 0.1 \cdot \mathbf{d}_i(t))$ has been added to each individual demand at each sample, where $\mathbf{d}_i(t)$ is the consumption of node i at sample t without noise.

Finally, the network model is simulated using EPANET in order to obtain pressures at the defined sensors and distribution of flows at the inputs. Base demands and boundary conditions (total flow and pressure set points) have been obtained from real measurements provided by the Barcelona water utility AGBAR. A random noise N(0,0.01mwc) has been added to pressure measurements after simulating the network.

Parameterization

The number of demand components and sensors used depends on both the final application of the calibration and the budget for installing sensors. This section considers five sensors in order to mimic the situation found in the real network, where five sensors were installed (Figure 6.11). These five sensors restrict the number of demand components that can be calibrated, as the system of equations in the well formulated calibration problem has to be over or equally determined. Consequently, the methodology presented in Chapter 3 will be used to define the memberships of nodes to five demand components, and the location of the five pressure sensors that are going to be used. Flow sensors will be considered in

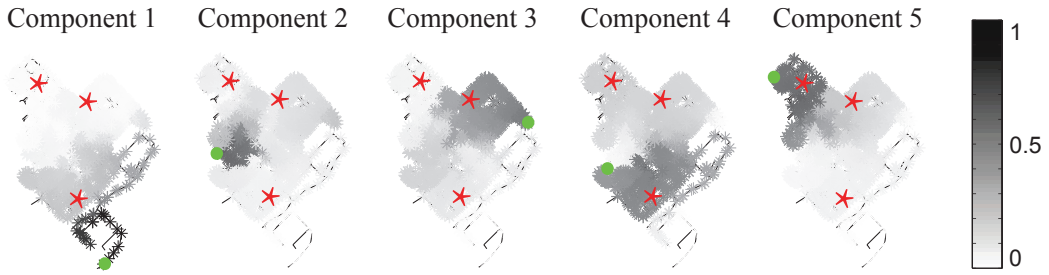

Figure 6.18: Memberships of nodes to five demand components in Nova Icària network with synthetic data. Each representation of the network depicts a grayscale map with the membership of each node to a particular demand component: the darker the node in the map, the higher the membership of the node to the demand component. The sensor with the highest sensitivity to variations in each demand component is also depicted in each map as a green circle. The three simulated leaks have been signaled with red stars

future studies. The parameterization and sampling design processes have to be performed again because the data from the synthetic reality generated is not the same as the real data in subsection 6.2.2.

Figure 6.18 depicts, in each of the network maps, the membership of each node to a particular demand component: the darker the node, the higher the membership to that component. Each map in Figure 6.18 also includes the location of the sensor (green dots) with the highest sensitivity to the component drawn.

Generation of scenarios

Nine leakage scenarios have been generated to evaluate the performance of the methodology developed. Leaks are assumed to be located at the nodes of the network. This simplification implies a loss of accuracy of the order of the pipe length. Such simplification can be assumed if the maximum localization error required by the company is greater than this length (Pérez et al. [2014]). In order to simulate a leak, an emitter coefficient C_e is set in a node so that the leak size generated depends on the pressure of that node (Rossman [2000]), as described in Equation 6.2.

$$q_L = C_e \cdot p^{\gamma} \tag{6.2}$$

where q_L is the leak water discharge, C_e is the emitter coefficient, p is the pressure at the node, and γ is an exponent of about 0.5 (Torricelli orifice equation).

Three different leak locations (signaled in Figure 6.18 with red stars) and

Table 6.9: Synthetic leakage scenarios generated, indicating the leak average water discharge, and the percentage leak consumption over the total consumed water

Leak		L1			L2			L3	
Scenario	S1	S2	S3	S4	S5	S6	S7	S8	S9
Mean daily water discharge	5l/s	3l/s	1l/s	5l/s	3l/s	1l/s	5l/s	3l/s	1l/s
% of total consumption	13%	8%	2.5%	13%	8%	2.5%	13%	8%	2.5%

three different sizes of leaks have been tested. Leak 1 (L1) is located in the effect zone of component $\mathbf{c}_5$; leak 2 (L2) is located in the effect zone of component $\mathbf{c}_3$; and leak 3 (L3) is located in the effect zone of component $\mathbf{c}_4$. Table 6.9 presents the main characteristics of the generated scenarios.

Results presented in the following section consider leaks appearing at low consumption hours ($\approx$ 20 l/s). Additional scenarios (not included in this work) where leaks occur at the peak consumption hour ($\approx$ 50 l/s) have been also tested, obtaining similar results.

Calibration results

The calibration process is applied considering the five components and sensors that have been selected. The values of the five demand components are calibrated by minimizing the error in pressure and flow measurements at each hour using the LS-based methodology presented in Chapter 4. Figure 6.19 depicts two weeks (without weekends) of calibrated demand component $\mathbf{c}_5$ values (white circles) and its 95% confidence intervals (gray boundaries). The first week (day 1 to 5) represents a non-faulty scenario. At the beginning of the second week (days 6 to 10), a 5 l/s leakage appears.

The validation of the calibrated components is done by comparing the average percentage of water consumption calculated from the calibrated values with the one calculated from billing. Figure 6.20 depicts this validation in two scenarios: a) No leakage scenario; and b) 5 l/s leakage scenario. Each of the radius represents a different demand component. Figure 6.20.a verifies the success of the calibration, whereas Figure 6.20.b warns of a bad calibration that has to be analyzed.

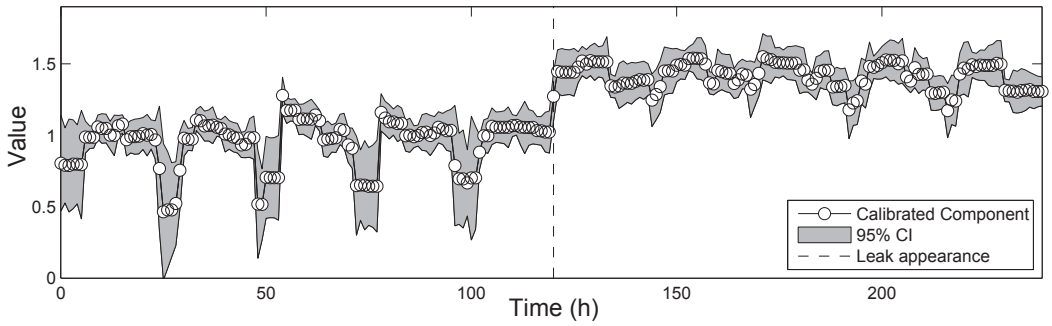

Figure 6.19: Calibrated demand component $\mathbf{c}_5$ during a non-faulty week and a faulty week with a 5 l/s leak. 95% confidence intervals (CI) are depicted in gray. The dashed line indicates the beginning of the leak

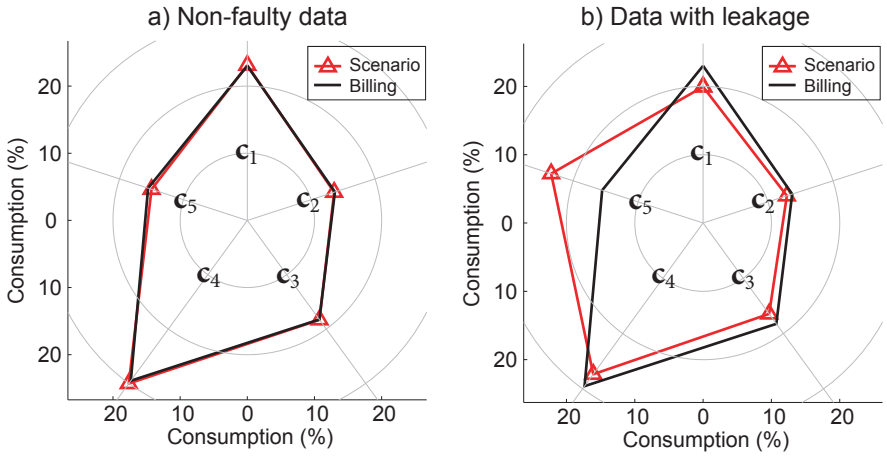

Figure 6.20: Average percentage of demand components water consumption from billing (black line) and calibrated components (red line with triangles) in scenarios: a) No leakage, and b) 5 l/s leakage in the area predominated by component 5

Selection of detection indicators' time windows and thresholds

The six detection indicators presented in Chapter 5 have to be particularized for the current case study. A time window of 12h is selected for the calculation of the detection indicators to detect changes in a fast but reliable way. However, the correlation and unit norms indicators have to be computed with a 24h time window due to their instability when calculated with a narrower window.

The selection of thresholds has to be done on a non-faulty state of the network. In this work, the non-faulty scenario is known (Figure 6.20.a). In a real case, the validation of the calibration presented in Figure 6.20 would be used to advise about the state of the network. In case of network experiencing undetectable burst or background leakage (Figure 6.20.b) before applying the methodology presented, this leakage would be considered as part of the demand model and thresholds would be set without taking it into account. The methodology would still be able to detect and locate new leaks occurring from that moment on, as explained in subsection 5.2.3.

Figure 6.21 shows the six indicators with the defined thresholds for each one. The 80% and 95% confidence intervals (CI) are marked with dashed and dash-dotted lines, respectively. These thresholds have been computed using the component with highest probability of having a false alarm during the non-faulty scenario in each of the detection indicators.

Figure 6.22 depicts the sum of scores obtained from the indicators during the non-faulty scenario. Only demand components $\mathbf{c}_1$, $\mathbf{c}_2$ and $\mathbf{c}_5$ get no null. The highest score is obtained by demand component $\mathbf{c}_1$ with a value of 3. Consequently, the global detection threshold is set at a value of 4 (dashed line in Figure 6.22).

Leak detection and localization

The methodology is tested using the nine faulty scenarios defined in Table 6.9 plus a non faulty scenario (S0). Table 6.10 sums up the results for all the scenarios in terms of detection, detection time and localization accuracy. Accuracy is presented as the distance (geographic and pipe distance) between the real leak and the node selected by the methodology as the one with highest probability to contain the leak. These distances are computed for both the *direct method* and the *leak membership method* presented in Section 5.3. The best result for each distance is highlighted in boldface letter.

Figure 6.23 depicts the graphical results for scenarios S3 (Figure 6.23.a,b), S4

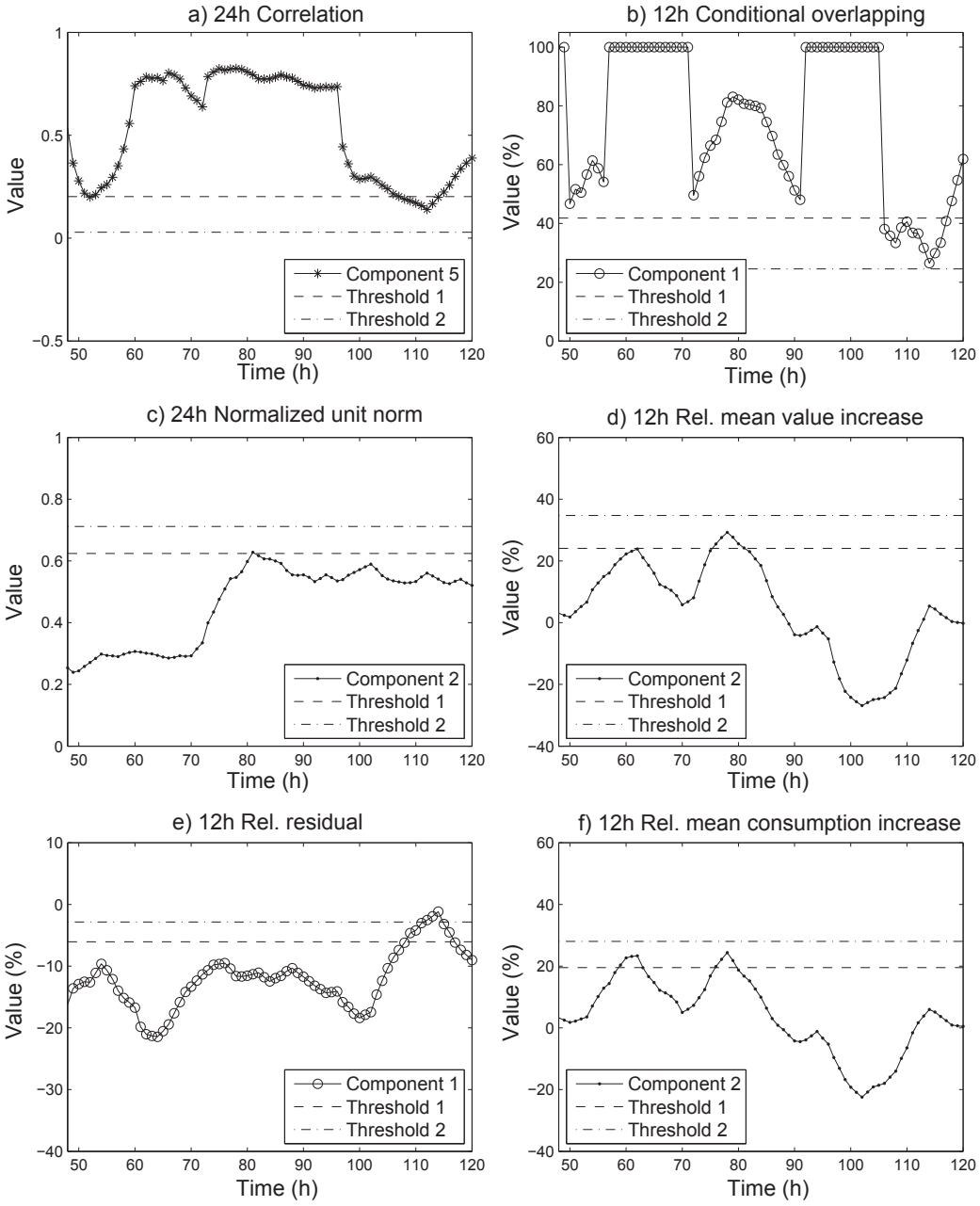

Figure 6.21: Detection indicators during the non-faulty scenario (training stage) with defined thresholds in dashed and dash-dotted lines for the demand component with worst value obtained in each indicator

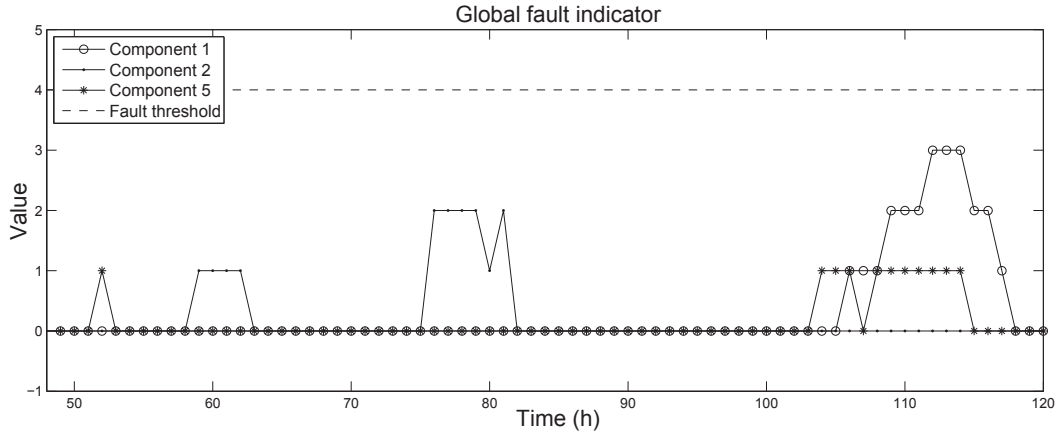

Figure 6.22: Sum of detection indicators scores for demand components 1, 2 and 5. The remaining demand components have all a null score. The global fault indicator threshold has been set at a value of 4 so that no alarm is triggered if the network remains in the non-faulty state

Table 6.10: Summary of results obtained in Nova Icària synthetic leak scenarios, indicating if leak is detected, the detection time, and the geographic and pipe distance to the real leak with the direct and leak membership methods. The best result for each scenario and distance is highlighted with boldface numbers

Scenario	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9
Leak detected	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Detection time	-	3h	4h	4h	4h	6h	6h	6h	10h	-
Geogr. distance to real leak [*direct*] (m)	-	**183**	183	**183**	657	657	657	220	220	-
Geogr. distance to real leak [*leak memb.*] (m)	-	224	**177**	**183**	**206**	**185**	**527**	**145**	**145**	-
Pipe distance to real leak [*direct*] (m)	-	**231**	**231**	**231**	857	857	857	365	365	-
Pipe distance to real leak [*leak memb.*] (m)	-	396	**231**	**231**	**293**	**263**	**698**	**181**	**181**	-

(Figure 6.23.c,d) and S8 (Figure 6.23.e,f) using greyscale maps. The first column of subfigures (Figure 6.23.a,c,e) refers to the *direct method*, whereas the second column (Figure 6.23.b,d,f) refers to the *leak membership method*. The darker the color in the greyscale map, the higher probability of the node to contain the leak.

Figure 6.24 depicts the geographical distance from the real leak to all nodes in the network (x axis), together with the indicator that gives a probability for the fault occurring in each node (y axis). For the direct approach (Figure 6.24.a,c,e), the indicator is the normalized membership; and for the leak membership approach (Figure 6.24.b,d,f), the indicator is the correlation. Each row of subfigures corresponds to scenarios S3 (Figure 6.24.a,b), S4 (Figure 6.24.c,d) and S8 (Figure 6.24.e,f). The node with the highest indicator value is shown with a red dashed line. Figure 6.25 depicts the same information but in terms of pipe distance from each node to the real leak. This distance helps to assess the use of acoustic methods that can locate precisely the leak if it is within a determined pipe distance. The teams looking for the leak would start from the node with highest probability of containing it (dashed line in Figure 6.25). The search direction is given by the leak probability of nodes in the vicinity of the one with highest probability.

Discussion

Leakage is detected in 8 out of the 9 faulty scenarios, as seen in Table 6.10. The 1 l/s leak located in demand component $\mathbf{c}_4$ (S9) is the only one that has not been detected. The high consumption of the component (≈30% of the total network consumption) masks the effect of the already low leakage water discharge (2.5% distributed among all components) and consequently, the changes in detection indicators are not large enough to identify a leak.

The non-faulty scenario is tested by considering a validation scenario (S0) with different boundary conditions than the one used to set the thresholds. A good result is obtained as no false alarms are triggered during this scenario.

All the evaluated leaks have been located in the component with highest memberships in the leak zone. Memberships are defined depending on the nodes' pressure sensitivity, thus any anomaly that affects pressure will have a greater impact on the predominant demand component of the anomalous zone than in any other demand component. This was the expected behavior that motivated the use of geographically distributed parameters to locate leaks.

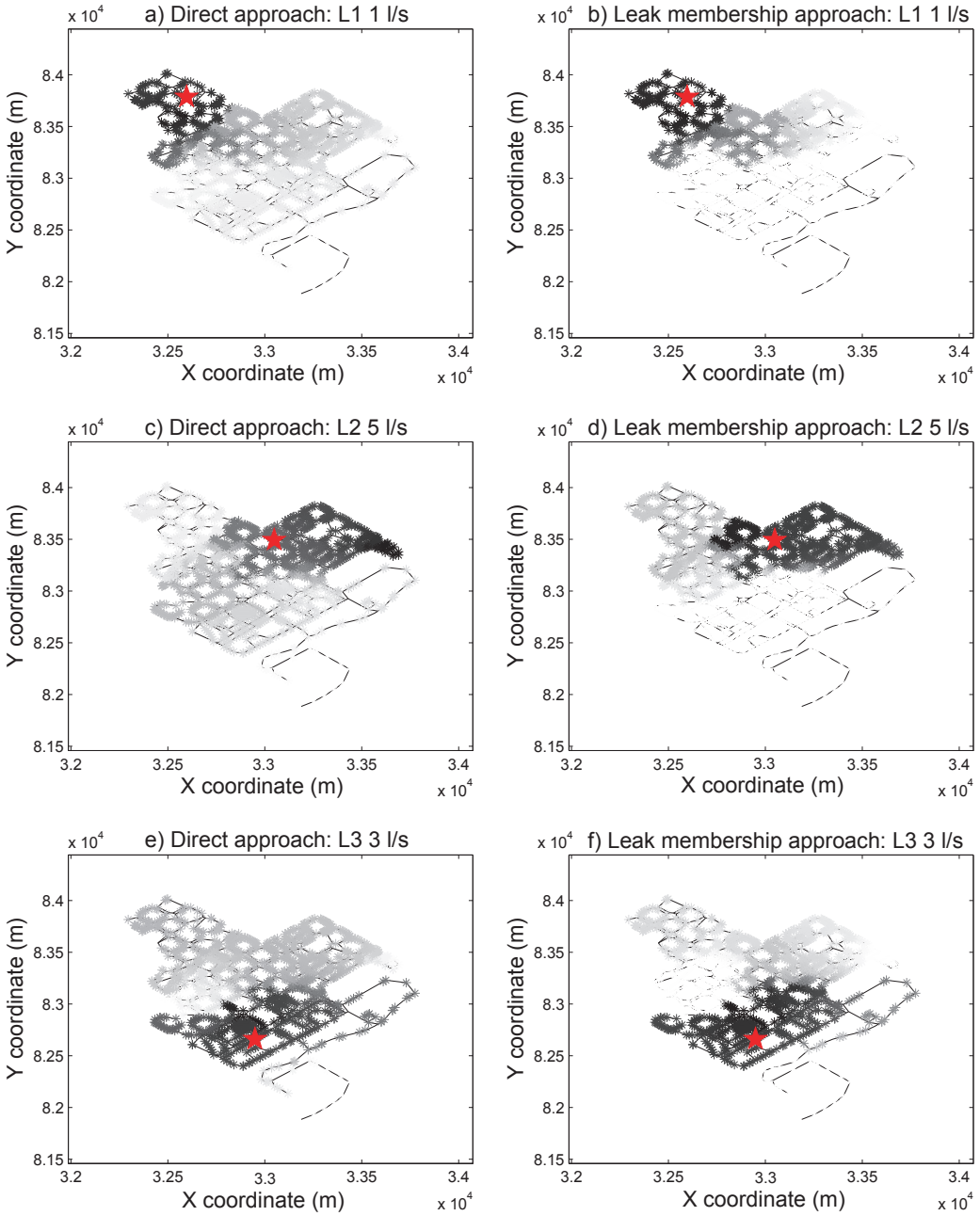

Figure 6.23: Localization results for scenarios S3 (first row of subfigures), S4 (second row of subfigures) and S8 (third row of subfigures). Column 1 presents gray scale maps with the membership of each node to the demand component that has triggered the alarm (*direct approach*), whereas column 2 presents gray scale maps with the correlation between each node memberships and the computed leak memberships (*leak membership approach*)

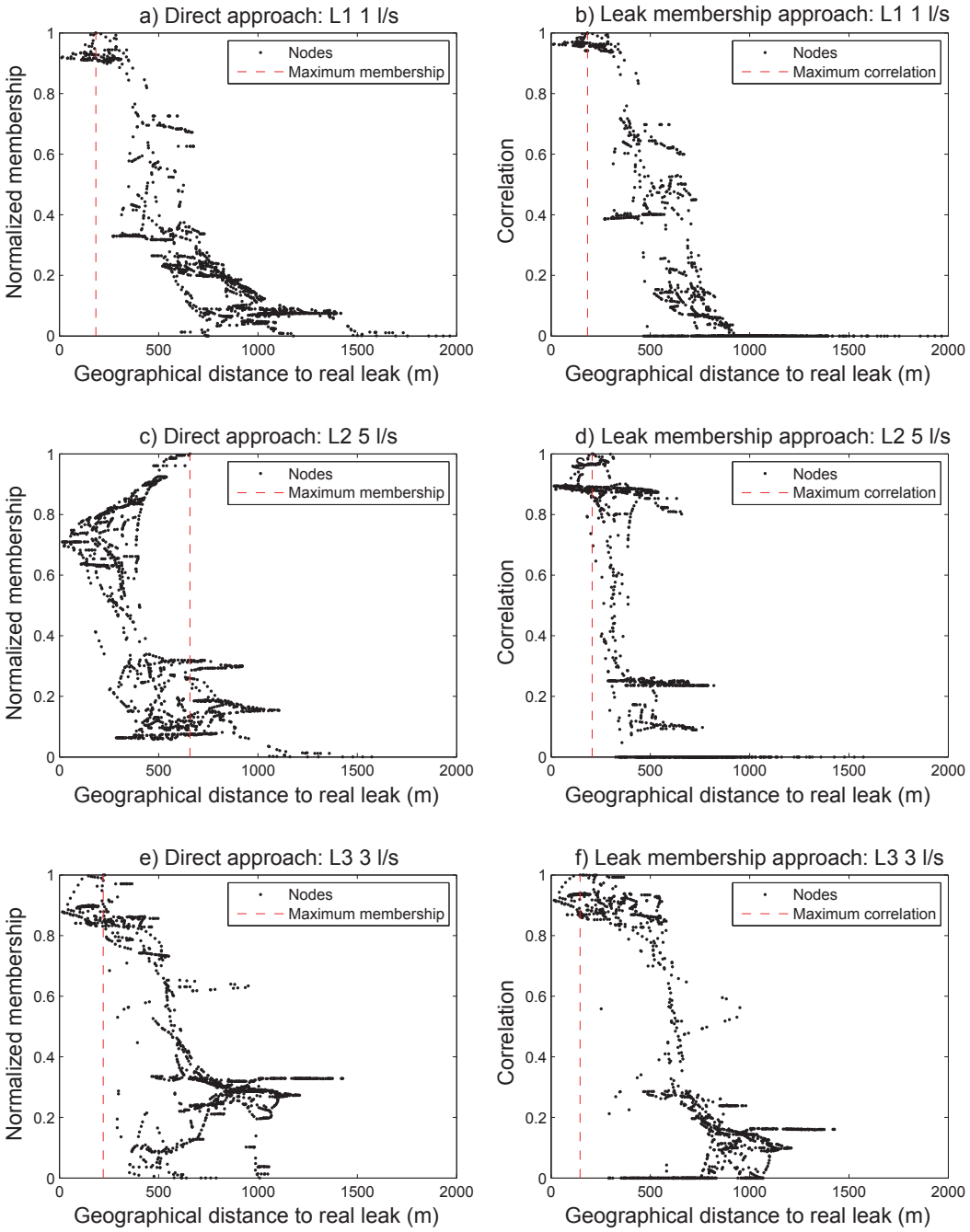

Figure 6.24: Geographical distance between the located leak and the real leak in scenarios S3 (first row of subfigures), S4 (second row of subfigures) and S8 (third row of subfigures). Subfigures in column 1 present results obtained with the direct approach. Every dot represents the distance from the node to the real leak depending on the normalized membership of that node to the demand component with highest score. Subfigures in column 2 present results obtained with the leak membership approach. Every dot represents the distance from the node to the real leak depending on the correlation between the node memberships and the computed leak memberships

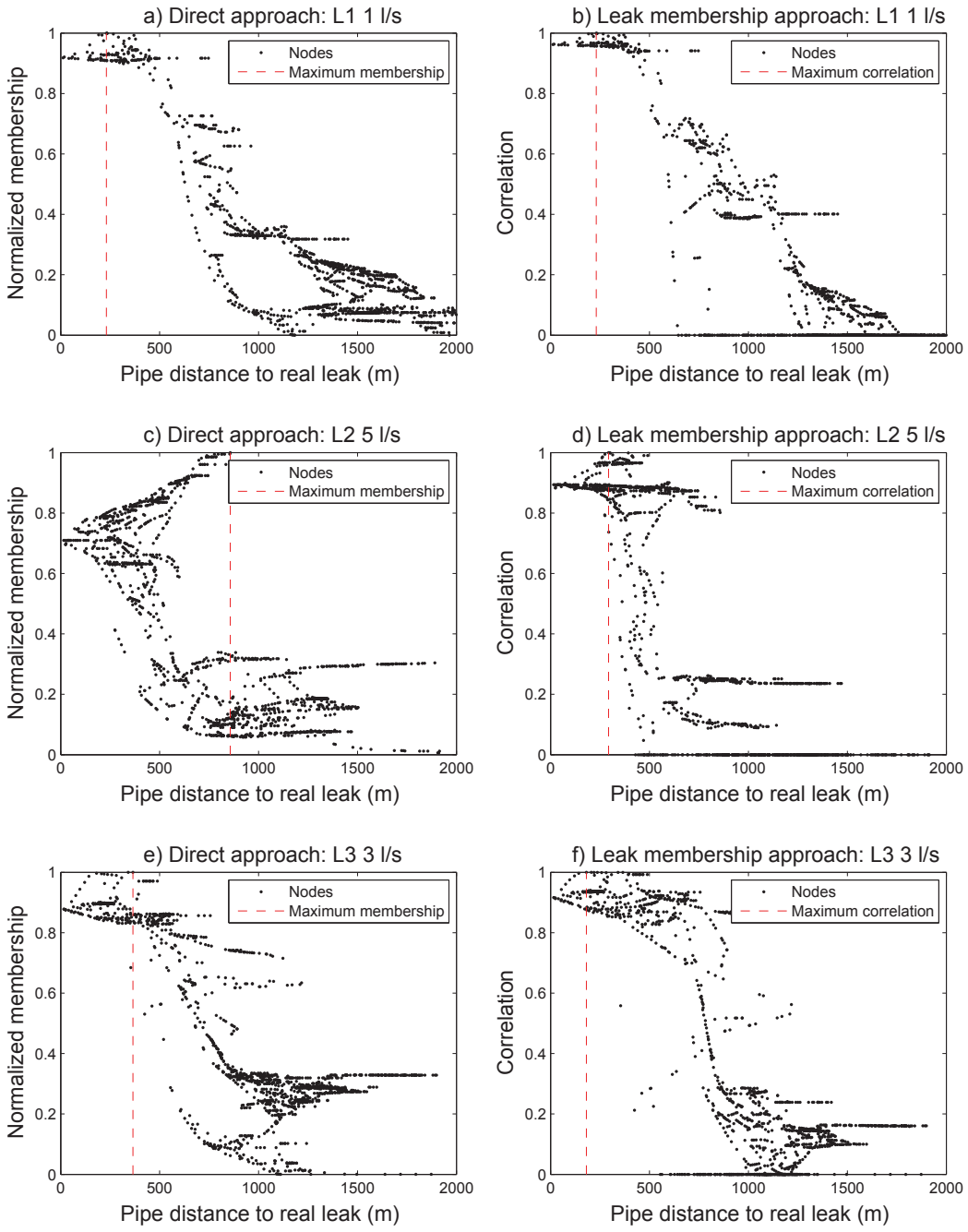

Figure 6.25: Pipe distance between the located leak and the real leak in scenarios S3 (first row of subfigures), S4 (second row of subfigures) and S8 (third row of subfigures). Subfigures in column 1 present results obtained with the direct approach. Every dot represents the distance from the node to the real leak depending on the normalized membership of that node to the demand component with highest score. Subfigures in column 2 present results obtained with the leak membership approach. Every dot represents the distance from the node to the real leak depending on the correlation between the node memberships and the computed leak memberships

Detection times depend on the relation between leak size and water consumption of the predominant demand component in the leak zone. This relation is directly linked to the variations in calibrated demand components: low consumption demand components are more affected by leaks than high consumption ones, in the same way that leaks with high water discharge have a greater effect than leaks with low water discharge. Hence, large variations in demand components are instantly identified by the detection indicators, whereas small variations require a larger number of time samples to be analyzed to identify if an anomaly is occurring or not.

The *direct method* only considers the effect of the leak on the demand component with higher nodes' memberships in the leak zone. If the affected demand component covers a large area, and the nodes with highest memberships to this component are far from the real leak, the localization accuracy could be decreased, as it happens in scenario S4. On the other hand, the *leak membership method* presents better results in terms of localization accuracy because it considers the effect of the leak on all demand components. The localization accuracy generated by the *leak membership method* is about 180 meters in all scenarios except in scenario S6. Pipe distances are greater than the geographic ones, but present an equivalent qualitative behavior in terms of accuracy, as seen in Figure 6.24 and Figure 6.25. The worst result is obtained for the 1 l/s leak 2 (S6) due to the small leak size together with its location in a zone where the predominant component has low memberships (30%-40%). The changes in the demand components are significant enough to detect the leak but not to locate it accurately.

To conclude, the methodology is able to distinguish between demand evolution and burst appearance. Daily, weekly and seasonal changes cannot be confused with leakage because: 1) calibrated demands are considered to have daily periodicity;, and 2) the comparison between demand components uses data from the same samples of the previous week. On the other hand, the long term evolution is progressively incorporated in the model by the continuous update of online calibrated demand components. This evolution is assumed to have slower impact on the online calibration than the one caused by a burst.

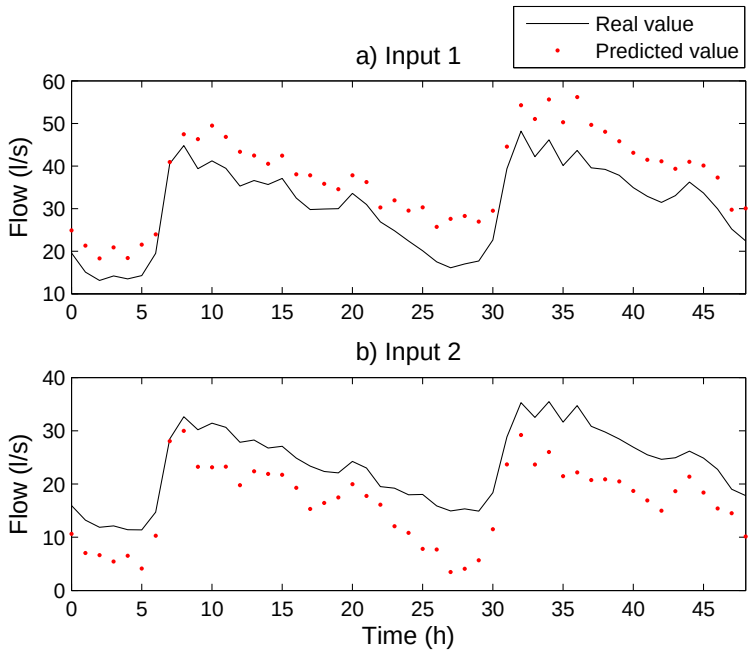

Figure 6.26: Nova Icària real (black line) and predicted (red dots) flows at the network inputs from December 19th 2012 to December 20th 2012

6.2.4 Leak detection and localization with real data

Finally, the third test consists in the application of the methodology presented in Chapter 5 and tested with synthetic data in subsection 6.2.3 to a real scenario, where a real leak was forced by the water utility.

Data analysis

A real leak of approximately 5,6 l/s was provoked in the Nova Icària WDN by the water utility on December 20th 2012 at 1 am. The only data available correspond to December 19th 2012 at 0 am to December 21st 2012 at 1 am, with the leak still persisting. Data from the week before the leak appearance (December 10th 2012 to December 14th 2012) are used to perform the depth correction and the parameterization process. These data have been already used in subsection 6.2.2. Consequently, the sensors' corrections are the ones detailed in Table 6.4.

Figure 6.26 shows the real and predicted flows values at each network input. Predicted data have been obtained by simulating the network model with the given boundary conditions using the basic demand model presented in Equation 2.12. Figure 6.27 shows the pressure sensors measurements for the same dates, and the pressure predictions using the basic demand model.

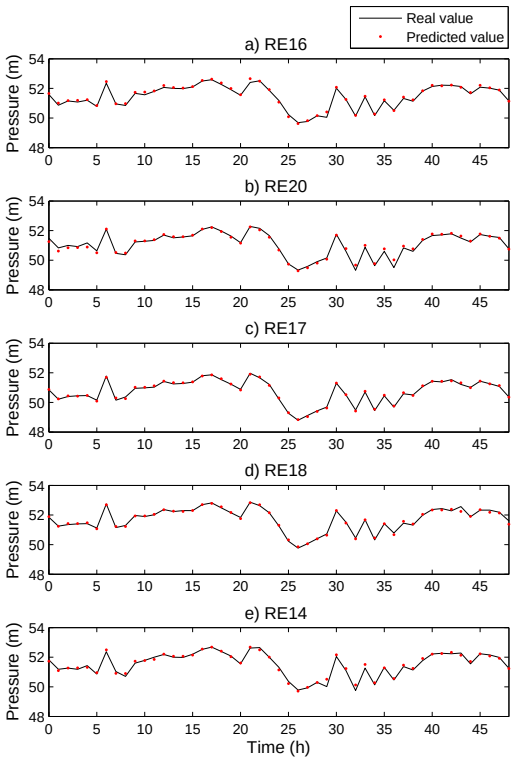

Figure 6.27: Nova Icària real (black line) and predicted (red dots) pressures at the five installed sensors from December 19th 2012 to December 20th 2012

Parameterization

In order to utilize the maximum amount of information available, five demand components are defined for the leak detection and location process. The addition of an extra demand component, compared to the four demand components defined in subsection 6.2.2, helps to reduce the size of the demand components, permitting a more accurate location of leaks, as the areas covered by each demand component become smaller.

The memberships of each nodal demand to five demand components are computed using Pseudo-code 1. Figure 6.28 depicts, in each of the network maps, the membership of each node to a particular demand component: the darker the node, the higher the membership to that component. Each map in Figure 6.28 also includes the location of the sensor with the highest sensitivity to the component drawn. Notice the relation between demand components in Figure 6.28 and Figure 6.13: Demand component $\mathbf{c}_2$ from the 4-component distribution is split into demand component $\mathbf{c}_2$ and $\mathbf{c}_3$ from the current 5-component distribution.

Table 6.11 sums up the average percentage of water consumption of each demand component: demand components $\mathbf{c}_2$, $\mathbf{c}_3$, and $\mathbf{c}_4$ have similar percentages

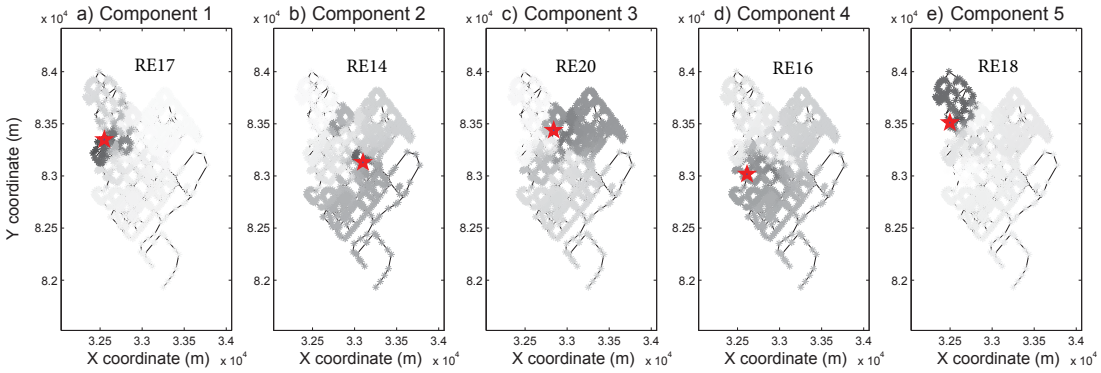

Figure 6.28: Memberships of nodes to each demand component in Nova Icària network considering the five available sensors. Each representation of the network depicts a grayscale map with the membership of each node to a particular demand component: the darker the node in the map, the higher the membership of the node to the demand component. The sensor with the highest sensitivity to variations in each demand component is also depicted in each map with a red star

Table 6.11: Average percentage of demand components' consumption in Nova Icària network computed from billing

Demand component	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\mathbf{c}_4$	$\mathbf{c}_5$
Average percentage of water consumption	10%	26.6%	23.5%	27.8%	12.1%

of consumptions (≈25%), whereas $\mathbf{c}_1$ and $\mathbf{c}_5$ have smaller consumptions (≈10%).

Online calibration

The online calibration process presented in Chapter 4 is applied to data from December 19th to December 20th. Table 6.12 collects the pressure prediction RMSE (first row), and the flow prediction RMSE (second row), for each demand model used. The percentage of error improvement has been also computed for each type of measurement, which represents the error improvement when using the demand components model instead of the basic demand model.

Figure 6.29 and Figure 6.30 depict the flow and pressure prediction errors for the same dataset. Each figure has two columns of subfigures, corresponding to the prediction error when using the basic demand model (1st column of subfigures); and the prediction error when using the demand components model calibrated using the LS method (2nd column of subfigures).

Finally, Figure 6.31 presents the average percentage of water consumption of each demand component, depending on the model and data used: the black

Table 6.12: Nova Icària pressure and flow prediction RMSE using the basic demand model and the demand components model. The percentage of improvement of the demand components model with respect to the basic demand model is also presented. Prediction errors have been computed with the calibration dataset

	Basic Model	**Demand Components Model**	
	RMSE	**RMSE**	**Improvement**
Pressure (m)	0.33	0.33	0%
Flow (l/s)	6.98	4.7	32.6%

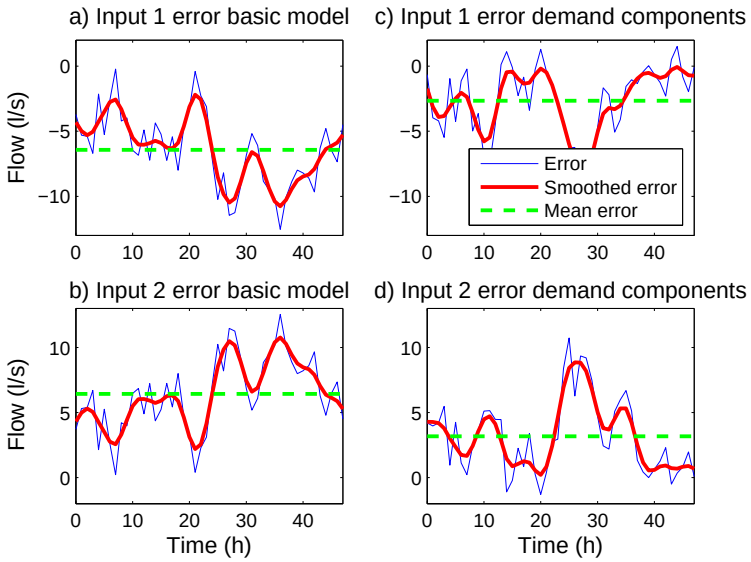

Figure 6.29: Nova Icària flow prediction error at each network input (rows of subfigures) using the basic demand model (first column of subfigures), and the demand components model (second column of subfigures). The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

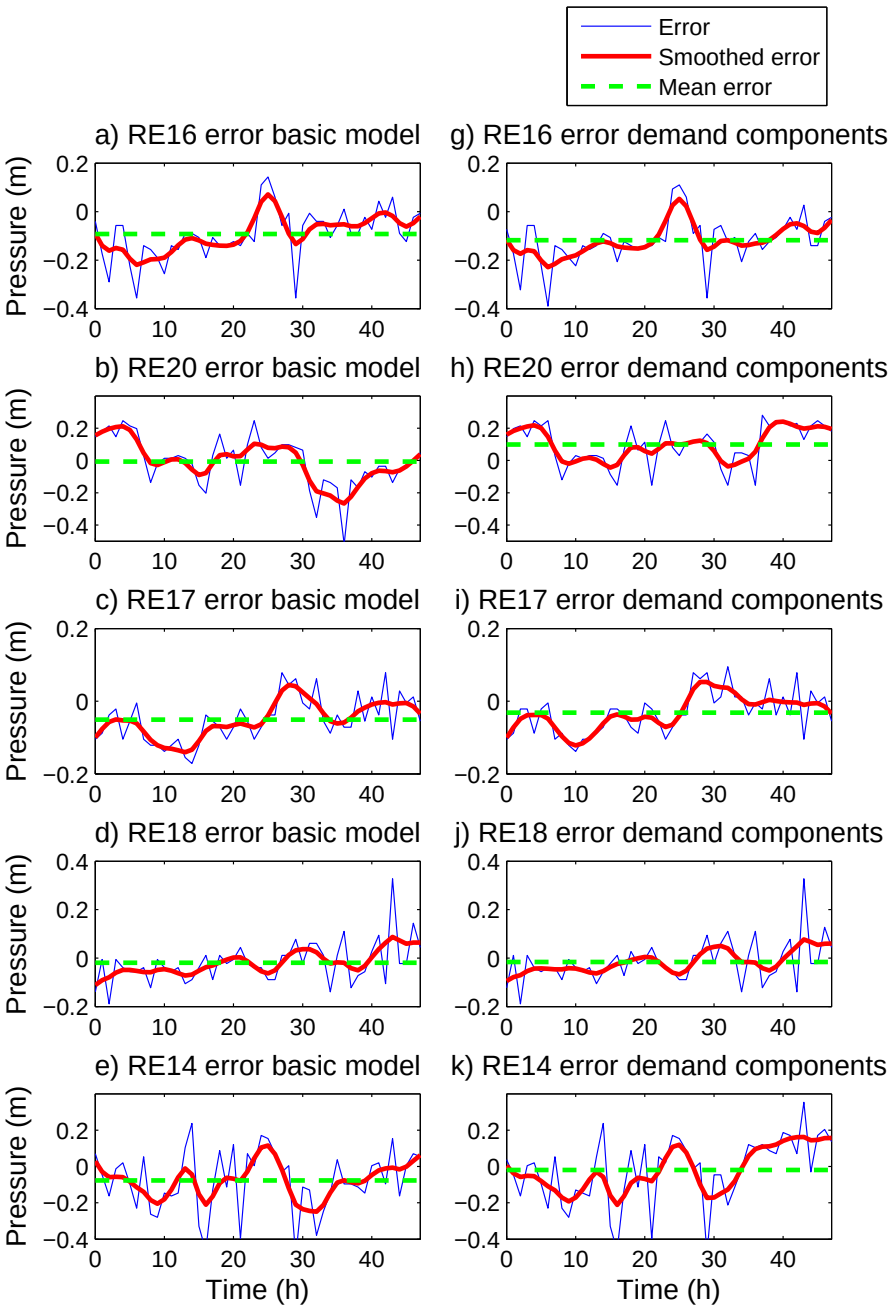

Figure 6.30: Nova Icària pressure prediction error at each sensor (rows of subfigures) using the basic demand model (first column of subfigures), and the demand components model (second column of subfigures). The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

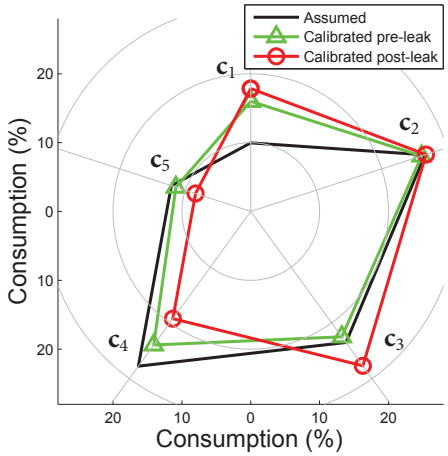

Figure 6.31: Average percentage of water consumption computed from billing (black line), LS-based demand components calibration before the leak appearance (green line with triangles), and LS-based demand components calibration after the leak appearance (red line with circles), in Nova Icària network

line corresponds to the basic demand model (assumed demands from billing), the green line with triangles corresponds to the demand components model calibrated with data before the leak appearance, and the red line with circles corresponds to the demand components model calibrated with data after the leak appearance.

Data generation

The limited data available impedes a direct application of the detection and localization methodology to test its validity. However, if we consider that in normal conditions the calibrated demand components during weekdays follow a similar behavior, we can enlarge the available dataset. This assumption is held by the results obtained in Canyars DMA (Section 6.1) and in Castelldefels Platja DMA (Section 6.3). The steps taken to generate the extra data are:

1. Apply the demand component calibration methodology explained in Chapter 4 and applied to this same network in subsection 6.2.2, considering $H_c = 1$ (only data from the current sample is used to calibrate the demand components), and five sensors/demand components.
2. Repeat four times the calibrated demand components from December 19th, where the leak was not still present.
3. Add a random Gaussian noise, with a standard deviation equal to a 20% of the current demand component value for all samples and repetitions. This percentage is obtained from two different sources: (1) In Canyars and

Castelldefels Platja DMAs, the demand components variations from one day to the following day is included in the 20% limits, and (2) the uncertainty during high consumption hours is again lower than 20% (17% in Canyars DMA and 10% in Castelldefels Platja DMA). Figure 6.32 depicts the calibrated demand components for the five days (four noisy demand components obtained from the calibration of December 19th, and the calibrated demand component for December 20th).

4. Use the first three days to define the detection indicators thresholds, and the global threshold.

5. Apply the online leak detection and location methodology to the real data from December 19th to December 20th.

Thresholds setting

The first three days from the new dataset are used to set the detection indicators thresholds, assuming that no leakage is present in the network. Background leakage, unknown status of valves and other errors will be masked by the thresholds, but new events occurring in the WDN will be detected if their effect makes the thresholds to be overtaken.

Figure 6.33 depicts, for the three first days of demand components available, the values obtained for each detection indicator, and the two detection thresholds set for each detection indicator following the methodology in Chapter 5. The values start at hour 48 because two days of data are required to start computing some of the indicators.

The highest score obtained during the non-faulty scenario corresponds to demand component $\mathbf{c}_4$ (green line with squares) at h=58..62, with a value of 2 (1 point of score from the relative mean value increase plus 1 point of score from the relative mean consumption increase). Consequently, the global threshold is set at a value of 3 to avoid false alarms during the non-faulty scenario.

Online detection

Once the thresholds have been set, data from December 19th to December 20th are used to evaluate the detection methodology. At each hour, the detection indicators are computed. If any of the computed detection indicators overtakes a threshold, the indicator gets a score depending on the threshold overtaken (1 point of score for the lower threshold, and 2 points of score for the higher

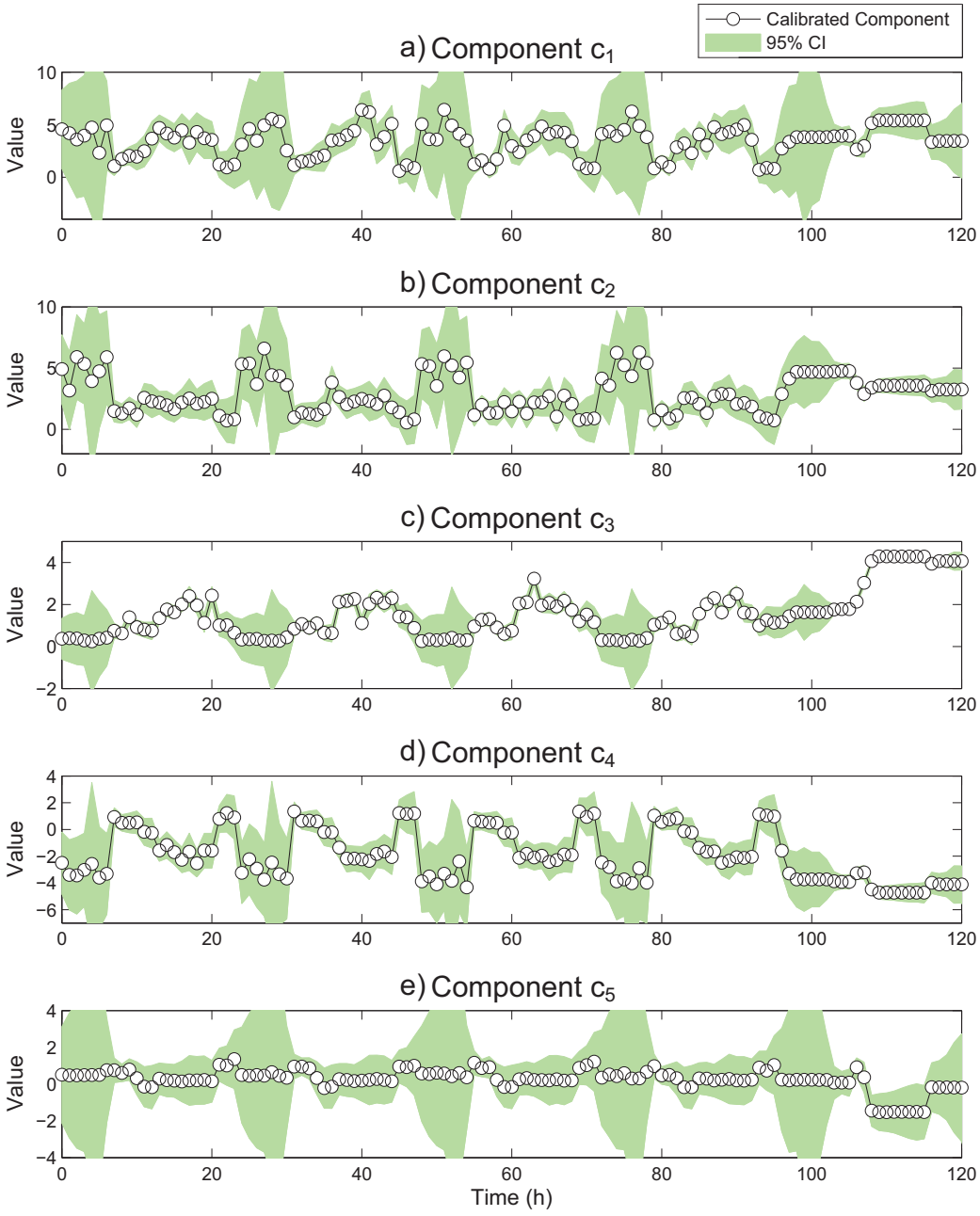

Figure 6.32: Nova Icària calibrated demand components with 95% confidence intervals (CI). Each subplot contains the noisy demand components from December 19th repeated four times, and the demand components from December 20th

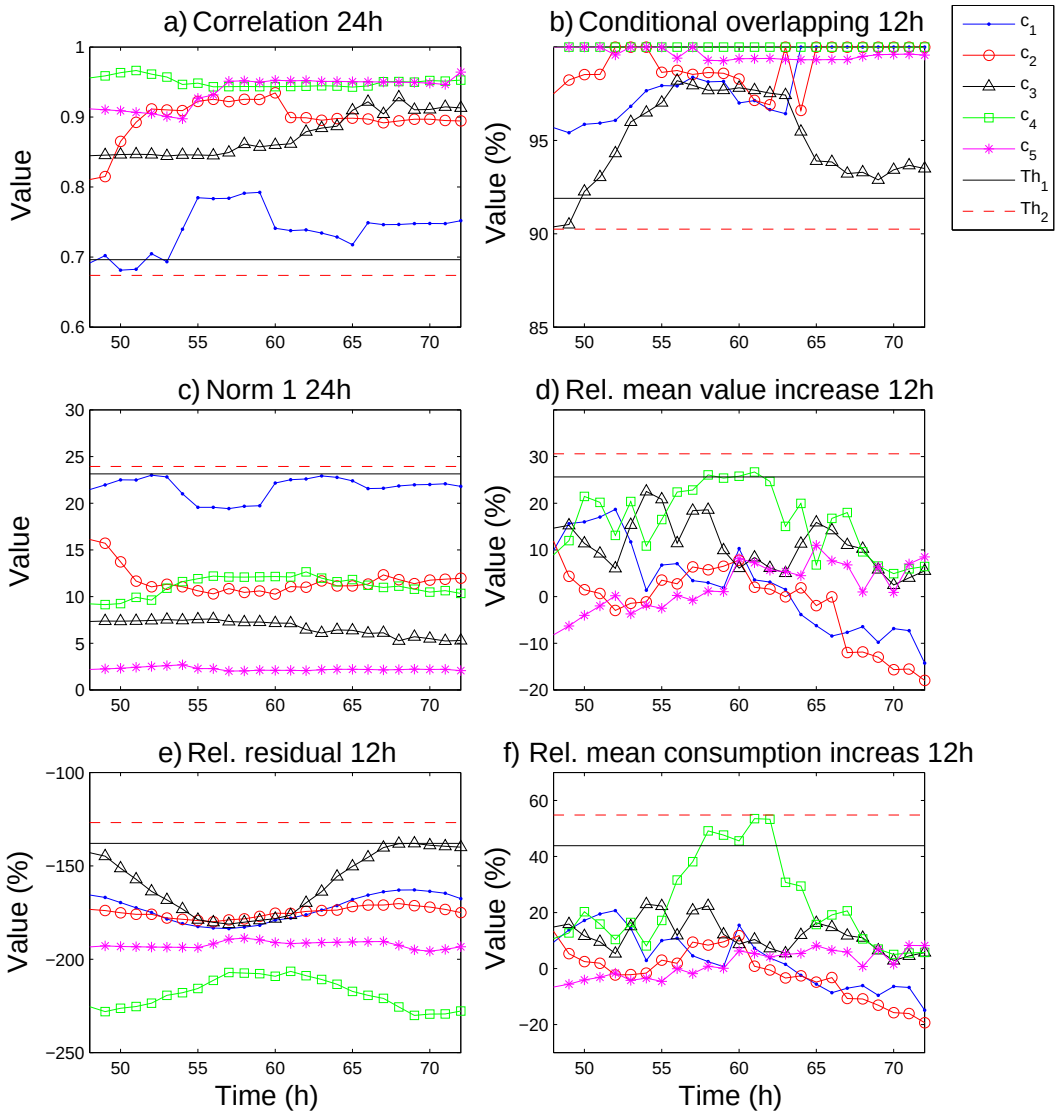

Figure 6.33: Nova Icària detection indicators values and set thresholds using the first three days of data (training phase)

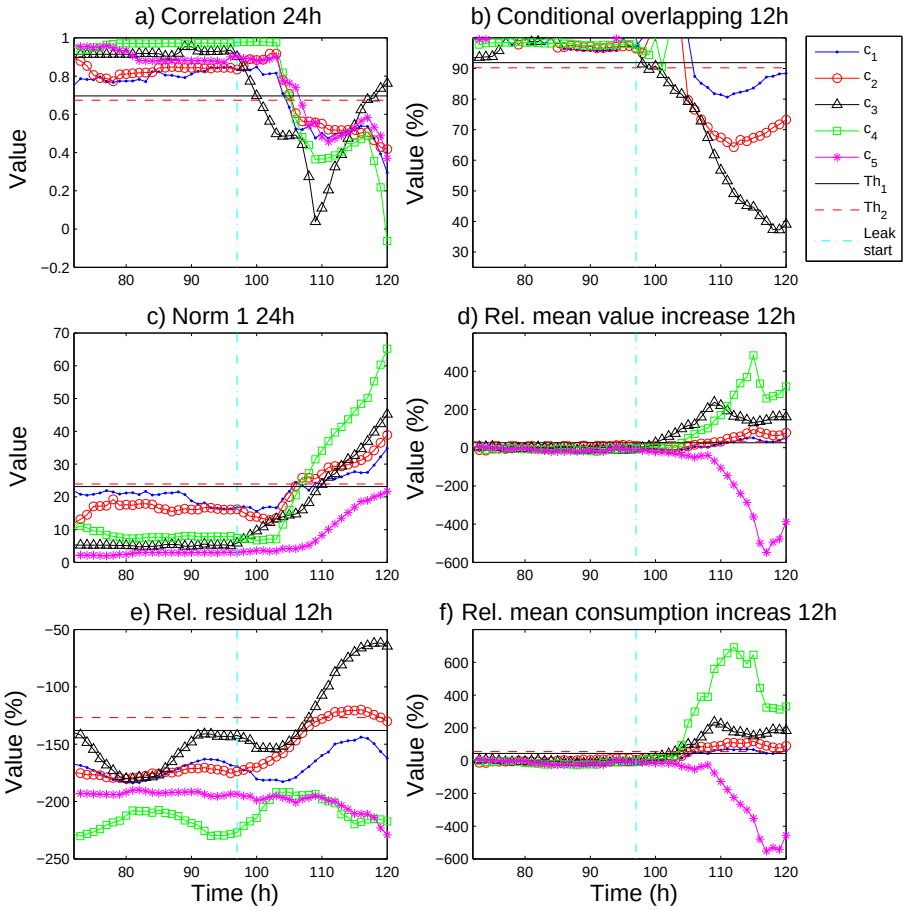

Figure 6.34: Detection indicators values and set thresholds for data from December 19th to December 20th. The leak starts at h=97

threshold). Figure 6.34 depicts the values for each detection indicator during the tested scenario. The cyan dash-dotted vertical line indicates the time when the leak starts (h=97).

Figure 6.35 presents the evaluation of each of the detection indicators presented in Figure 6.34 for the demand component $\mathbf{c}_3$. This demand component is the first component to trigger the global detection alarm, four hours after the leak appearance.

Online localization

Once the leak has been detected in the WDN, the leak localization process explained in Chapter 5 starts. Figure 6.36 presents the geographical distances from the real leak to the node selected as the one with the leak, for the direct (subfigures a), b), and c)) and leak membership (subfigures d), e), f)) approaches. Each row of subfigures presents a different localization time: The first row corresponds

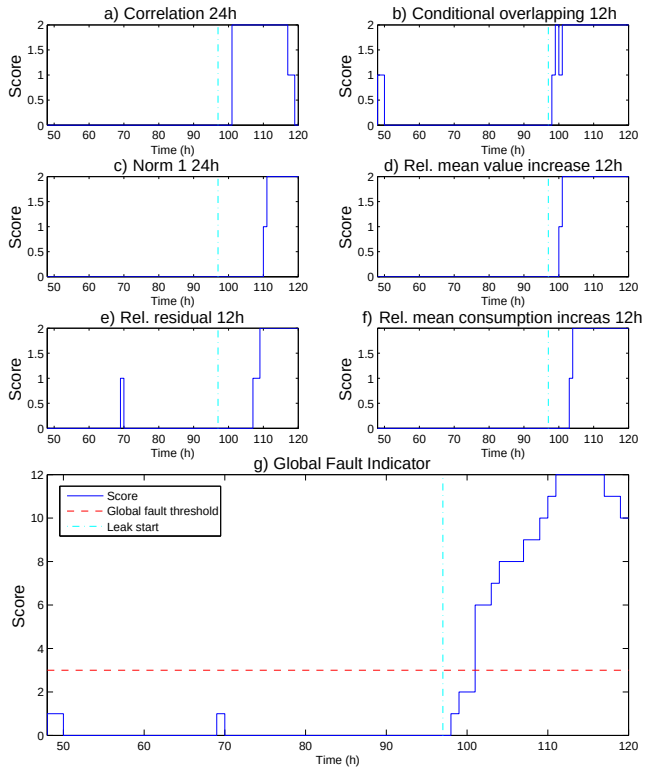

Figure 6.35: Evaluation of detection indicators for demand component $\mathbf{c}_3$

to the first hour after detecting the leak, the second row corresponds to 13 hours after the leak appearance (not detection), and the third row corresponds to 23 hours after the leak appearance. The distance errors between the signalled node and the real leak are: a) 167 m, b) 163 m, c) 163 m, d) 171 m), e) 184 m, and f) 474 m.

Finally, Figure 6.37 presents the localization map corresponding to the leak membership approach applied at the detection time (the map corresponds to Figure 6.36.d).

Discussion

The application of the leak detection and localization methodology in a real scenario has generated good results: leak is detected only four hours after its appearance (Figure 6.35), and located at the same time with an accuracy of 167 meters (Figure 6.36). The localization error is lower than the assumable error fixed by the water utility, which is around 200 meters. This fact allows the water utility to search the signaled zone for leak using acoustic methods.

Initially, the online calibration process has been applied, with a 33% improvement in the prediction of flow, and the same RMSE in pressure prediction as the

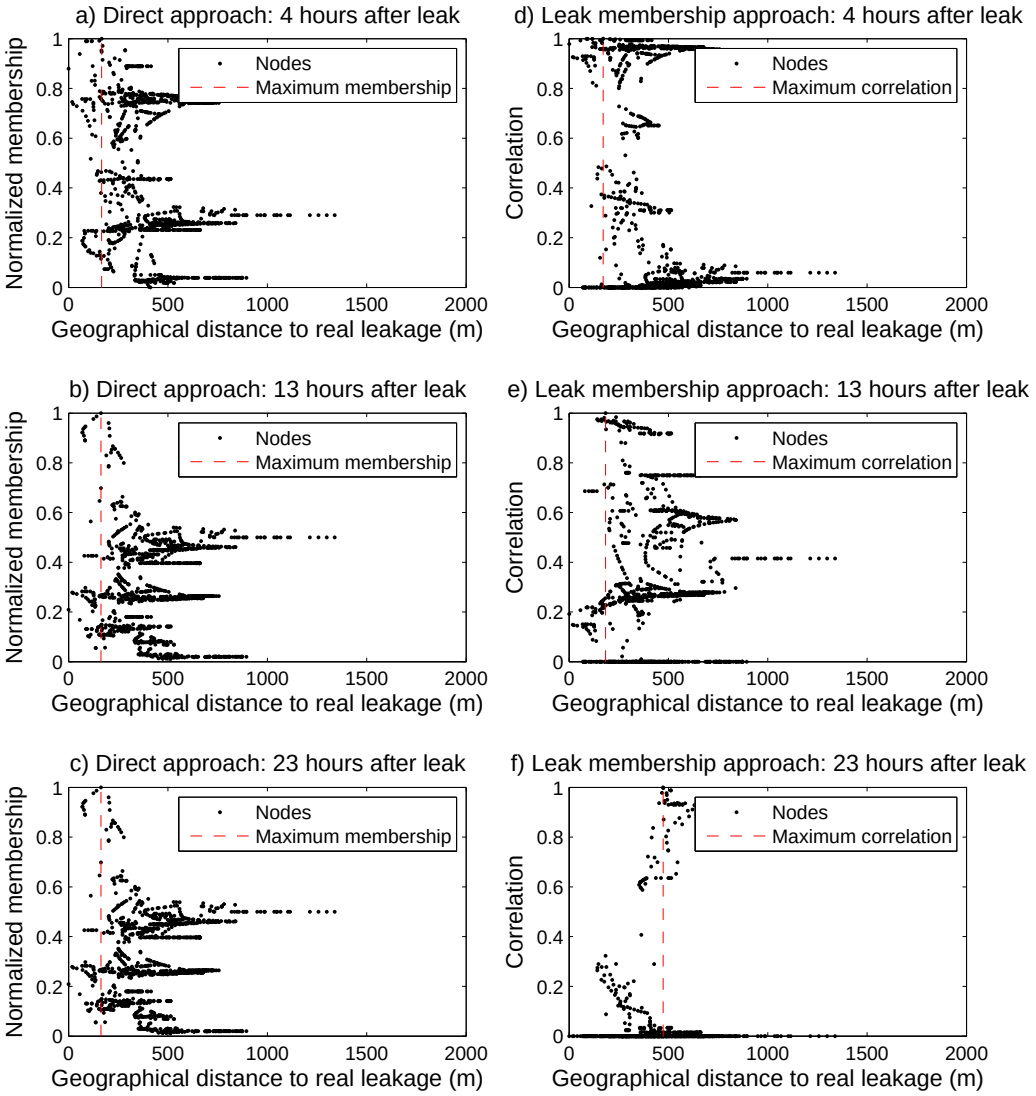

Figure 6.36: Geographical distance between the located leak and the real leak computed 4 hours after the leak appearance, when the leak is detected (first row of subfigures), 13 hours after the leak appearance (second row of subfigures) and 23 hours after the leak appearance (third row of subfigures). Subfigures in column 1 present results obtained with the direct approach. Every dot represents the distance from the node to the real leak depending on the normalized membership of that node to the demand component with highest score. Subfigures in column 2 present results obtained with the leak membership approach. Every dot represents the distance from the node to the real leak depending on the correlation between the node memberships and the computed leak memberships

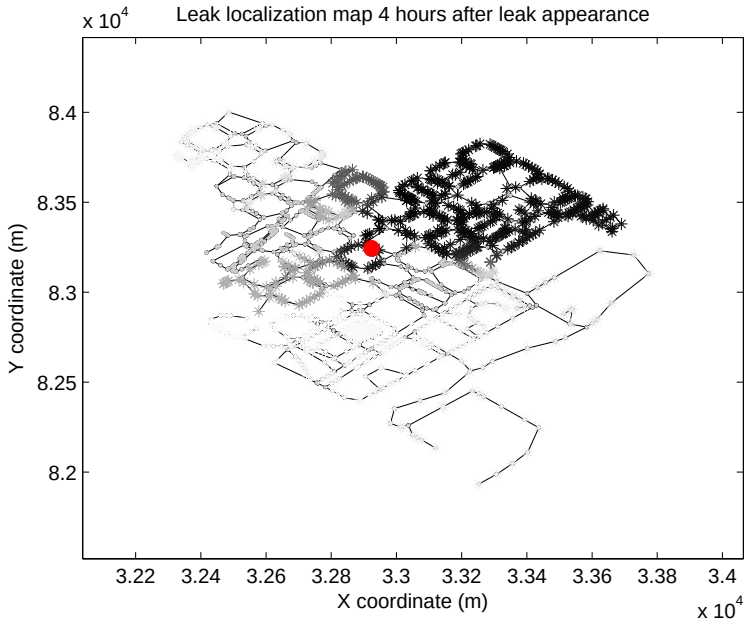

Figure 6.37: Gray scale map with the correlation between each node memberships and the computed leak memberships obtained after leak detection, i.e. four hours after the beginning of the leak. The real leak is signaled with a red circle

one obtained when using the basic demand model (Table 6.12). The generation of extra data using the non-leak day helps to overcome the lack of data, which will not happen in a future online-software.

During the calibration process, an analysis of Figure 6.31 would give a first suggestion of the leak appearance, as demand component $\mathbf{c}_3$ abruptly increases its consumption from 22% to 28%.

The detection stage is continuously performed together with the online calibration. Four hours after the leak appearance, the detection alarm is triggered (Figure 6.35), and the location method starts. During the initial 15 hours, both direct and leak membership approaches locate the leak with an accuracy of 163-184 meters, with the lower distance error obtained by the direct approach (Figure 6.36). The high localization error in Figure 6.36.f (474 m) has a direct explanation: the leak membership is computed from the demand components variation with a time window of 12 hours, thus after 12 hours of the leak appearance, the components' variations are not due to the leak anymore.

The leak detection and localization method improves the results from the location method presented in Pérez et al. [2014], where the same scenario was tested with a residual-based methodology, locating the leak 14h after its appearance, with similar location accuracy.

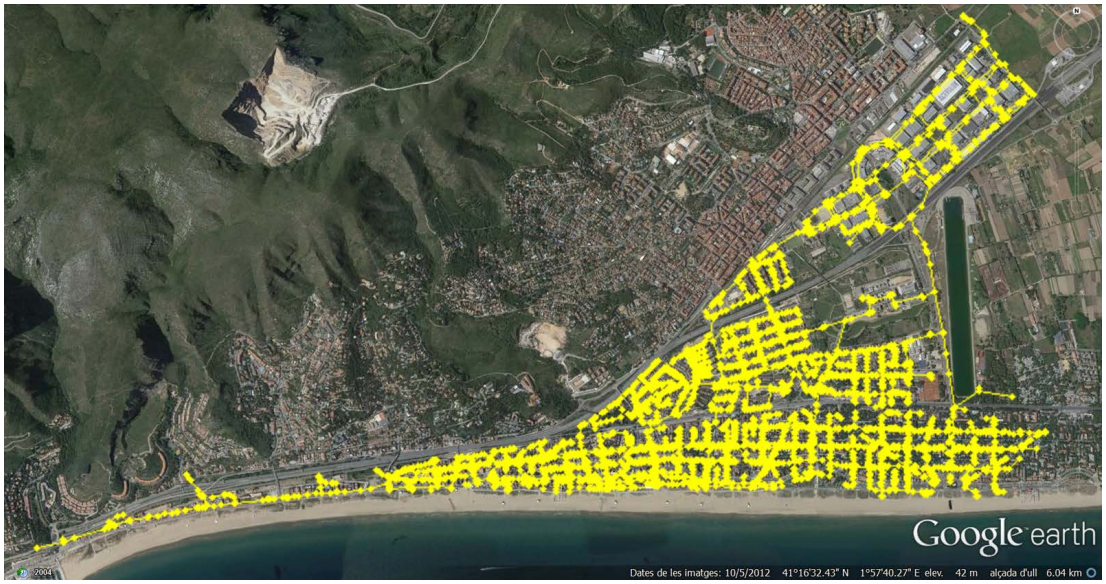

Figure 6.38: Castelldefels Platja DMA location in the catalan city of Castelldefels. Satellite image extracted from Google Earth

6.3 Standarization: Application to Castelldefels Platja network

Castelldefels platja DMA is situated in Castelldefels (Catalonia) (Figure 6.38), supplying water to nine neighborhoods: Baixador, Lluminetes, la Pineda, les Botigues, Can Bou, Mar-i-Sol, Gran Via Mar, Zona Universitària, and Camí Ral. It is composed of 5153 pipes and 4991 junctions. Water is supplied from the transport network through two pressure reduction valves. Pressure and flow are monitored at both water inlets with a sample time of 10 min. The resolution is 0.3 l/s for the flow sensors, and 0.1 mwc for the pressure sensors. The minimum night flow is of about 45 l/s, and the peak-hour flow is 85 l/s. A pressure control is applied to this network, fixing the pressure at the PRVs outlets to 40 meters during night-time and 50 meters during day-time. The average daily maximum head-loss between any two network junctions is 2 meters.

6.3.1 Context of the work

This last example is presented to show the straightforward application of the developed methodologies to any network. For this reason, each of the presented sections will include the amount of time and model simulations runs required to prepare the data and network model for the application of the methodologies,

Table 6.13: Engineering time (h), computational time (h), and number of EPANET simulation runs for the skeletonization process

Process	Engineer time (h)	Computational time (h)	EPANET runs
Skeletonization	0.01	0.22	144

and for the application itself. The test has been performed on a Dell laptop with the following characteristics:

- Windows 8.1 Pro x64
- Intel Core i7-4510U CPU 2.00GHz
- 8GB RAM
- Matlab 7.8.0 (R2009a) x32
- EPANET Toolkit 2.00.11 x32

6.3.2 Skeletonization

The huge size of Castelldefels Platja network claims for a model reduction process to avoid memory errors while decreasing the computational time required for the application of the methodology. The skeletonization process presented in Appendix A is applied to the network model, considering only those reductions that do not affect the hydraulic behavior of the network. Figure 6.39 depicts the original model and the reduced model. The number of junctions has been reduced from 4991 to 2794 (56%), and the number of pipes from 5153 to 2956 (57%). The reduced model has been evaluated by comparing the pressures in the 2794 remaining nodes with the same nodes in the original network model. The highest pressure error during 144 samples (one day, with a sampling time of 10 minutes) is 0.00033 meters, significantly lower than the sensors' resolutions (0.1 m). Table 6.13 sums up the time required for the skeletonization process.

6.3.3 Sampling Design

The parameterization and sampling design processes were performed to propose the location of six pressure sensors and the parameter definition for demand calibration, as explained in Chapter 3. However, the proposed sensors' locations differ from the final ones, which have been obtained from the methodology in Bonada et al. [2014] (based on leak detection). The installed sensors can still

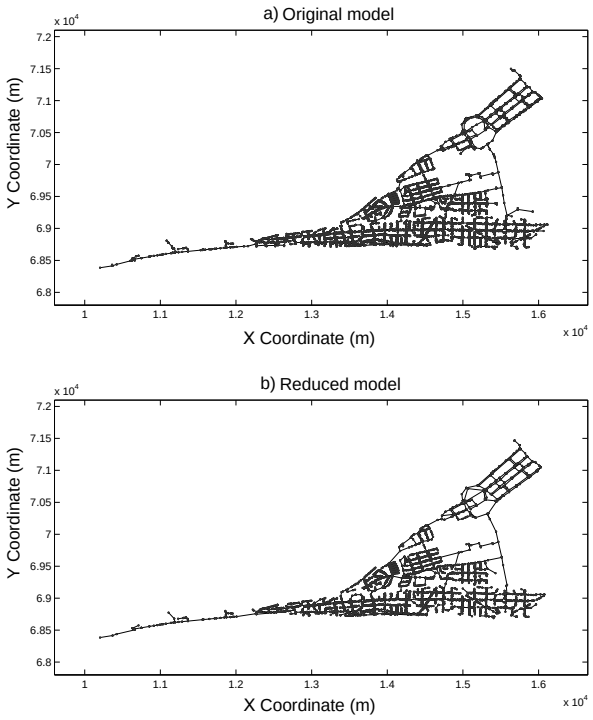

Figure 6.39: Castelldefels Platja network original and reduced EPANET models

Table 6.14: Castelldefels Platja sensors sampling times (minutes), and offset corrections (meters)

Sensor ID	*RE35*	*RE36*	*RE37*	*RE38*	*RE39*	*RE40*
Sampling Time (min)	10	30	10	10	10	10
Offset correction (m)	0.93	1.2	1.38	1.78	0.47	-0.38

be used to calibrate demands by defining the demand components depending on the available sensors' locations, thanks to the versatility of the proposed method (Section 3.3). Figure 6.40 depicts the proposed sensors locations with green circles, and the final locations with red stars. The resolution of the installed sensors is 0.1 mwc and the sampling times are defined in Table 6.14.

The time required for the sampling design process, which includes the computation of the sensitivity matrices and the parameterization process, is presented in Table 6.15.

6.3.4 Data Analysis

Data from May 25th 2015 (Monday) to May 31st 2015 (Sunday) are used to calibrate the demand components for a whole week. Data from June 1st 2015

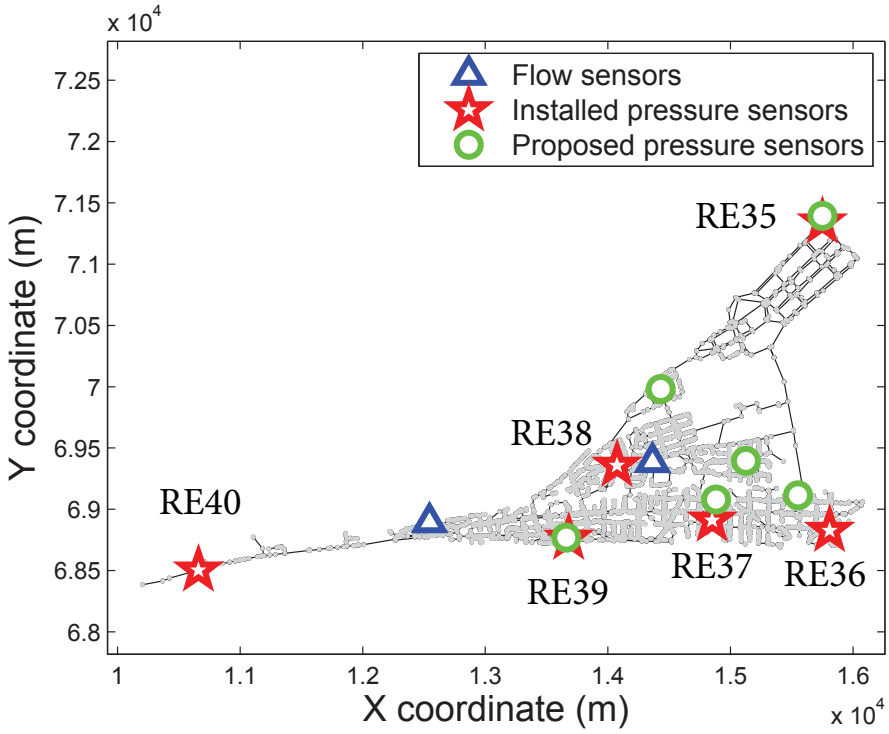

Figure 6.40: Castelldefels Platja EPANET network model with highlighted sensors. The network water inputs are signalled with a blue triangle; the installed pressure sensors are signalled with red stars and the proposed pressure sensors are signalled with green circles. The flow sensors are installed at the input pressure reduction valves, so that the total flow consumed in the network is known

Table 6.15: Engineering time (h), computational time (h), and number of EPANET simulation runs for the sampling design process

Process	**Engineer time (h)**	**Computational time (h)**	**EPANET runs**
Sensitivity calculation	0.01	0.15	144
Parameterization	0.01	1	0
Sampling design	0.01	0.01	0

(Monday) to June 2nd 2015 (Tuesday) are used to validate the demand components obtained. Previously, data from May 21st 2015 (Thursday) to May 24th 2015 (Sunday) are used to analyze and correct the data coming from the network, and to perform the parameterization process before the calibration starts. These three datasets will be referred as precalibration dataset, calibration dataset, and validation dataset. The validation dataset consists of only two days due to the limited access to real data.

Figure 6.41 shows the complete set of boundary conditions (input valves' pressure set points and flows) for the precalibration and calibration datasets. Figure 6.41.b) and c) show the real and predicted flow measurements at input 1 and 2, respectively. Predicted data have been obtained by simulating the network model with the given boundary conditions using the basic demand model presented in Equation 2.12. Figure 6.41.a) shows the aforementioned pressure control at the DMA inputs.

Figure 6.42 shows the real and predicted pressures at the six installed sensors for the precalibation and calibration dataset. The corresponding pressure prediction error for the precalibration dataset (4 days) is depicted in Figure 6.43. The blue line corresponds to the raw error using all data, and the red line represents the smoothed error, which has been computed by means of a smoothing spline. The green dashed line corresponds to the mean pressure prediction error. This error is treated as an offset that cannot be associated to the demand model. As suggested in Pérez et al. [2014], the offset is corrected to eliminate possible depths errors, model nodes' elevations inaccuracies, roughness modeling errors, or badly calibrated sensors' offsets. The same correction is also considered when using data from the calibration week. Table 6.14 contains the specific correction for each sensor.

Figure 6.44 shows the flow prediction error at the two network inputs when using the basic demand model during the precalibration dataset. It can be noticed that the error is positive or negative depending on the time of the day. Input 1 provides more water than expected during daytime, while input 2 provides more water than expected during nighttime. The demand components model should be able of distributing demands (in time and space) in the network to minimize these errors.

Table 6.16 contains the time and model simulation runs required for the analysis and correction of data.

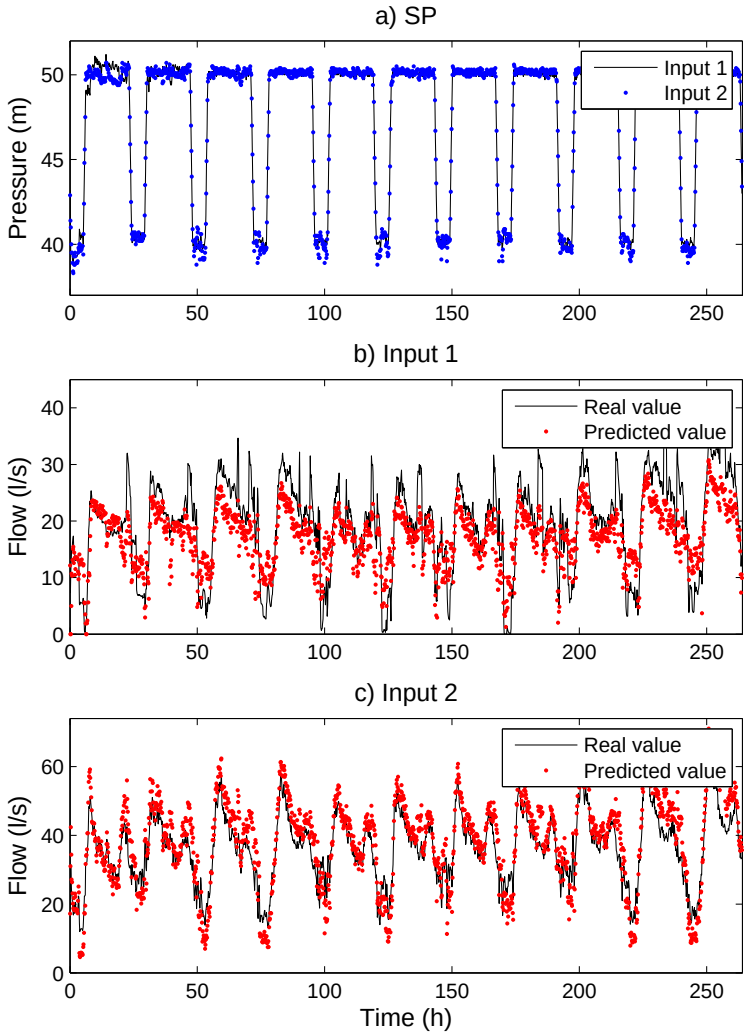

Figure 6.41: Castelldefels Platja boundary conditions from May 21st 2015 to May 31st 2015: a) Input pressure reduction valves set points, b) Real (black lines) and predicted (red dots) flow metered at input 1, and c) Real (black lines) and predicted (red dots) flow metered at input 2

Table 6.16: Engineering time (h), computational time (h), and number of EPANET simulation runs for the data analysis

Process	**Engineer time (h)**	**Computational time (h)**	**EPANET runs**
Data analysis	1	0.2	1584

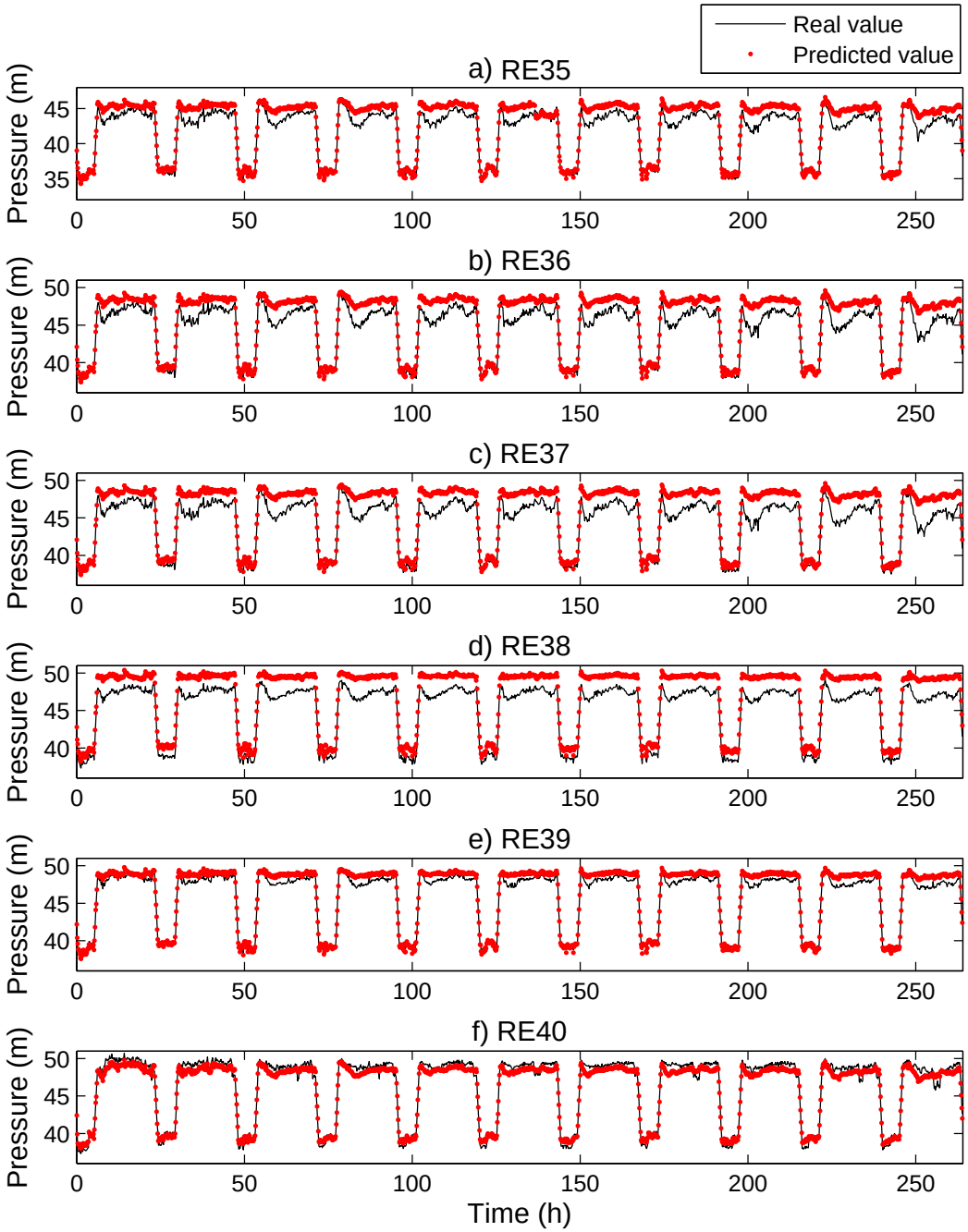

Figure 6.42: Castelldefels Platja real (black lines) and predicted (red dots) pressures at the six installed sensors from May 21st 2015 (Thursday) to May 31st 2015 (Sunday)

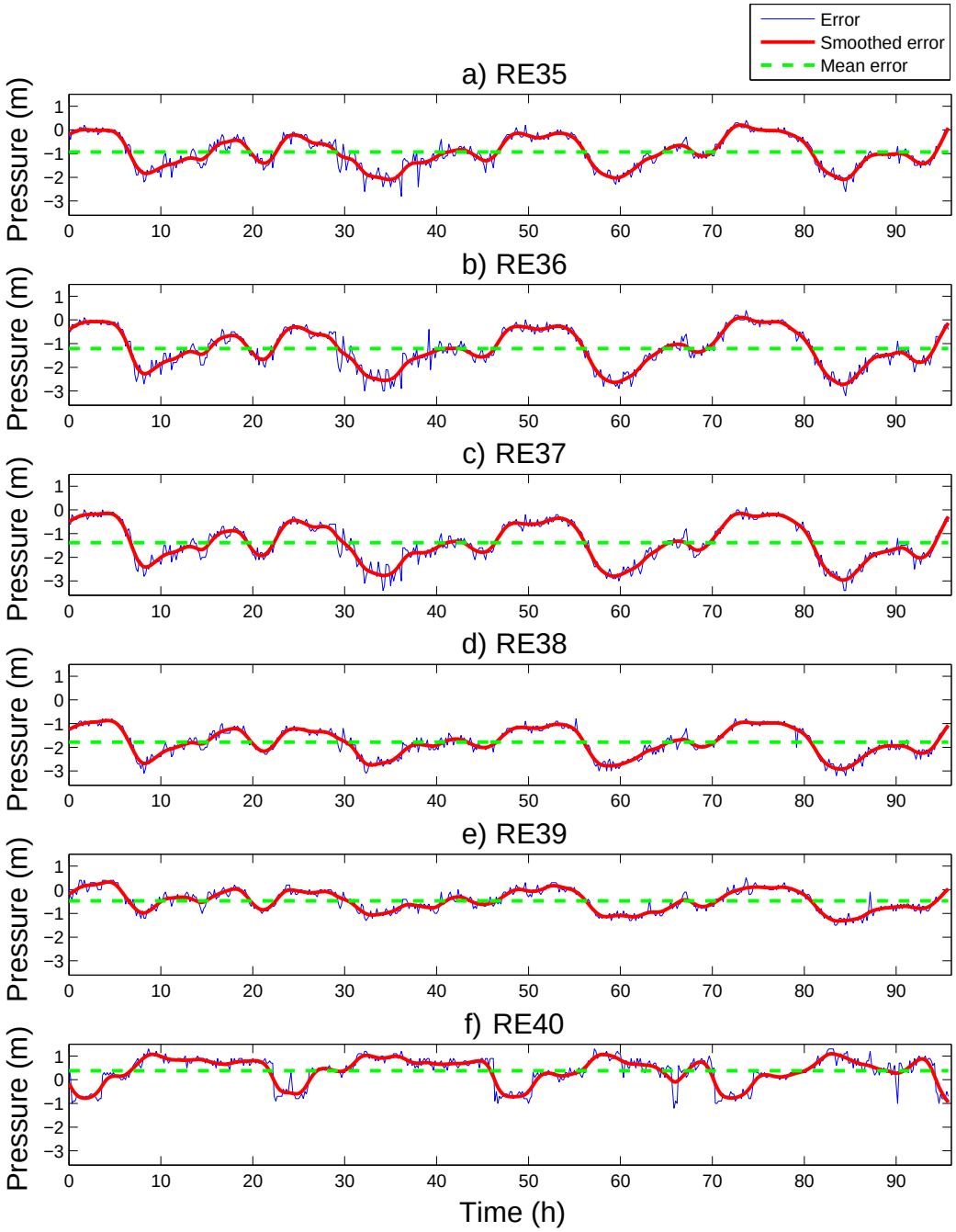

Figure 6.43: Castelldefels Platja pressure prediction error in the six installed sensors for the precalibration dataset. The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

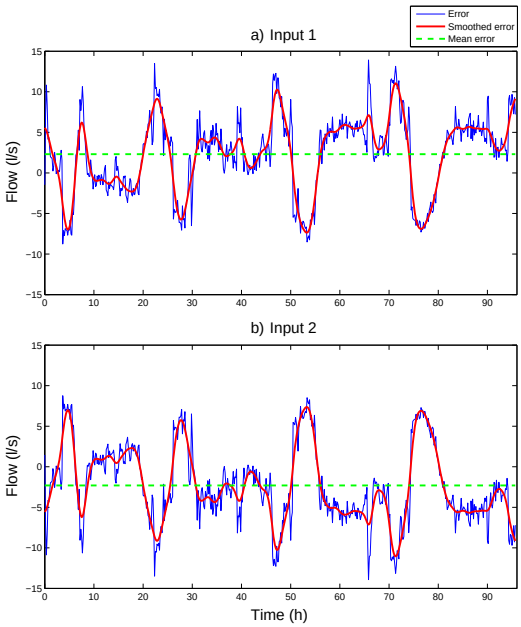

Figure 6.44: Castelldefels Platja flow prediction error at the network inputs for the precalibration dataset. The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

6.3.5 Parameterization

Data from the precalibration dataset are used to compute the sensitivity matrices to perform the parameterization process. The memberships of each nodal demand to six demand components are computed using Pseudo-code 1, considering the six installed sensors. Figure 6.45 depicts, in each of the network maps, the membership of each node to a particular demand component: the darker the node, the higher the membership to that component. Each map in Figure 6.45 also includes the location of the sensor with the highest sensitivity to the component drawn.

Table 6.17 sums up the average percentage of consumption of each demand component: demand component $\mathbf{c}_2$ has the lowest percentage of consumption (4%), whereas $\mathbf{c}_6$ consumes 35% of the water in the network. This information is used to analyze the calibration results: errors in the average percentage of consumption of the calibrated demand components compared to the assumed consumption in Table 6.17 can be assigned to background leakage, burst, fraudulent consumptions, unknown status valves, non-metered users, or wrong billing information.

The time required for the parameterization process, including the computation

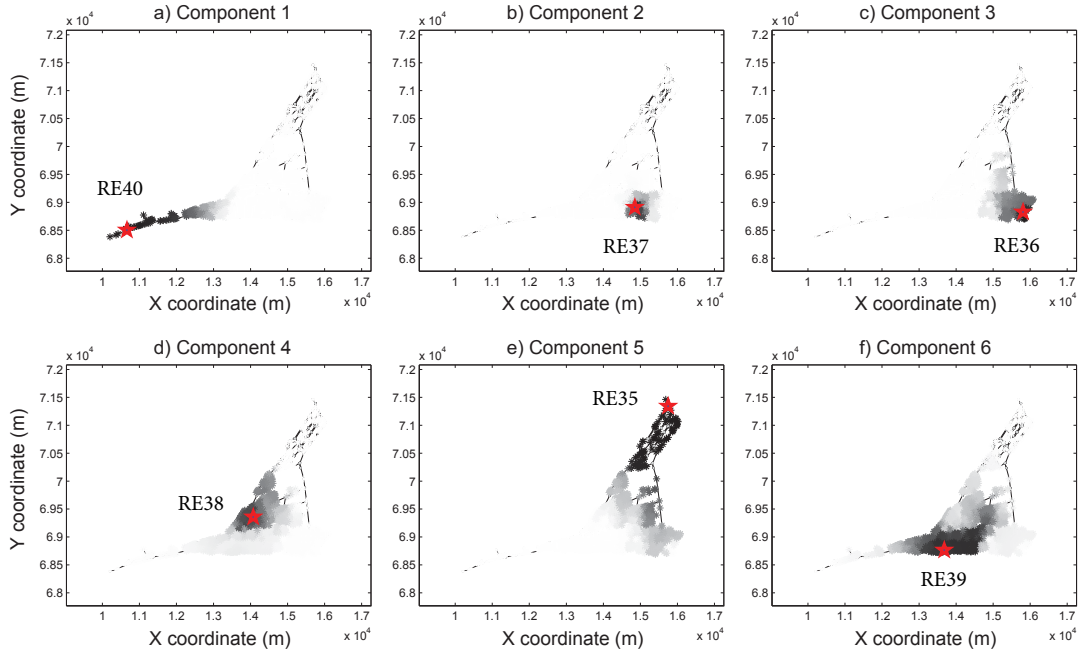

Figure 6.45: Memberships of nodes to each demand component in Castelldefels Platja network considering the six installed sensors. Each representation of the network depicts a grayscale map with the membership of each node to a particular demand component: the darker the node in the map, the higher the membership of the node to the demand component. The sensor with the highest sensitivity to variations in each demand component is also depicted in each map

Table 6.17: Average percentage of demand components' water consumption in Castelldefels Platja network computed from billing

Demand component	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\mathbf{c}_4$	$\mathbf{c}_5$	$\mathbf{c}_6$
Average percentage of water consumption	16,75%	4,04%	8,66%	13,96%	21,12%	35,47%

Table 6.18: Engineering time (h), computational time (h), and number of EPANET simulation runs for the parameterization process

Process	Engineer time (h)	Computational time (h)	EPANET runs
Sensitivity calculation	0.01	0.15	144
Parameterization	0.01	0.003	0

Table 6.19: Castelldefels Platja pressure and flow prediction RMSE using the basic demand model and the demand components model. The percentage of improvement of the demand components model with respect to the basic demand model is also presented. Prediction errors have been computed with the calibration dataset

	Basic Model	Demand Components Model	
	RMSE	RMSE	Improvement
Pressure (m)	0.788	0.472	40.1%
Flow (l/s)	5.339	1.596	70.1%

of the sensitivity matrices, is presented in Table 6.18.

6.3.6 Calibration

The calibration process presented in Chapter 4 has been applied to Castelldefels Platja network, considering the six pressure sensors and the defined demand components. A calibration horizon $H_c = 1$ has been used, i.e. data from the same hour of each day have been used to calibrate the demand component values for that specific hour and day, minimizing the prediction error only for that hour. The six samples within the same hour (sampling time of 10 minutes) have been filtered to reduce noise. The optimization method used is based on least squares minimization using the SVD to compute the inverse of the sensitivity matrix.

Figure 6.46 presents the calibrated demand components values for the calibration dataset (7 days). Demand component values are presented as white circles surrounded by the 95% confidence intervals (green boundaries).

Table 6.19 collects the pressure prediction root mean squared error (RMSE) (first row), and the flow prediction RMSE (second row), for each demand model used. The percentage of error improvement has been also computed for each type of measurement, which represents the error improvement when using the demand components model instead of the basic demand model. Table 6.20 collects the pressure prediction RMSE (first row), and the flow prediction RMSE (second

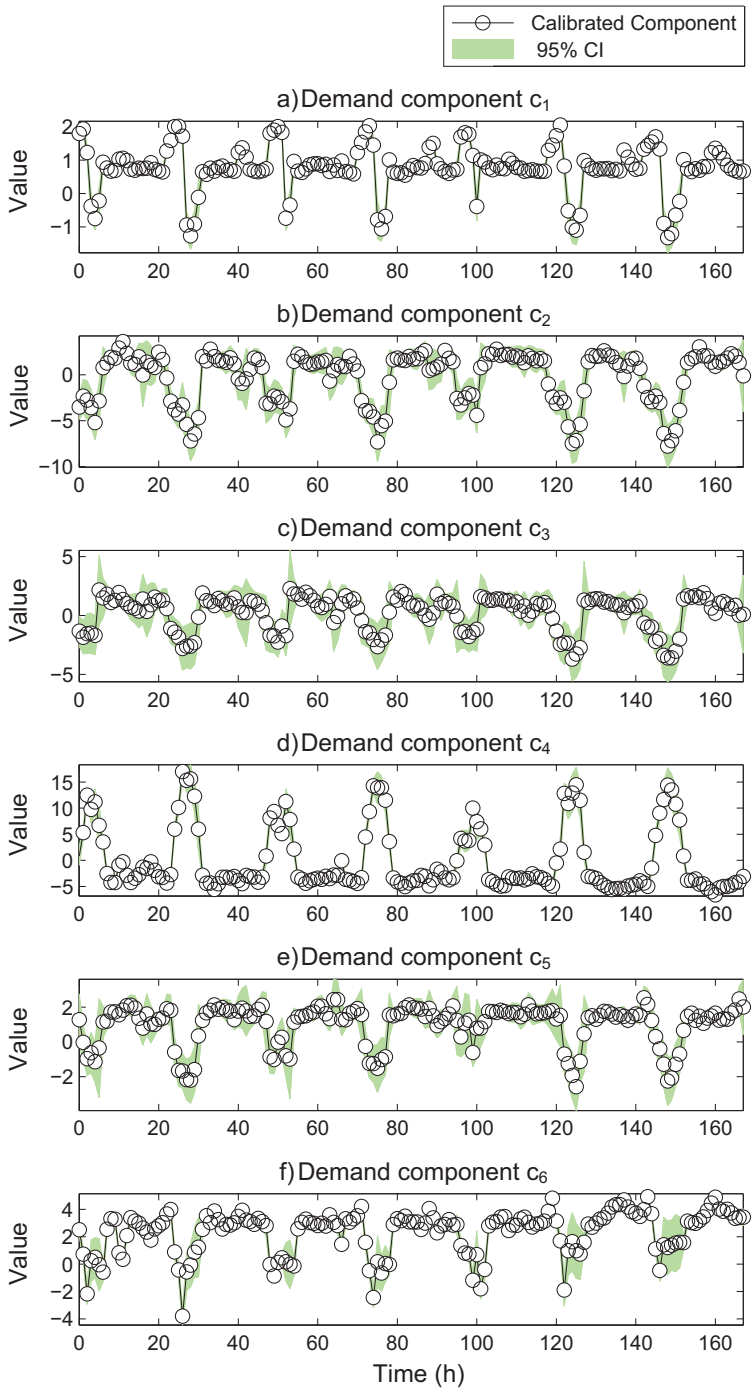

Figure 6.46: Castelldefels Platja calibrated demand components with 95% confidence intervals (CI) using LS optimization

Table 6.20: Castelldefels Platja pressure and flow prediction RMSE using the basic demand model and the demand components model. The percentage of improvement of the demand components model with respect to the basic demand model is also presented. Prediction errors have been computed with the validation dataset

	Basic Model	**Demand Components Model**	
	RMSE	**RMSE**	**Improvement**
Pressure (m)	0.865	0.616	28.7 %
Flow (l/s)	6.113	5.153	15.7 %

Table 6.21: Engineering time (h), computational time (h), and number of EPANET simulation runs for the calibration process

Process	**Engineer time (h)**	**Computational time (h)**	**EPANET runs**
Calibration	0.01	51	50400

row), for the validation dataset (two days).

Figure 6.47 and Figure 6.48 depict the flow and pressure prediction errors for the calibration dataset. Each figure has two columns of subfigures, corresponding to the prediction error when using the basic demand model (1st column of subfigures), and the prediction error when using the demand components model (2nd column of subfigures).

Finally, Figure 6.49 presents the average percentage of consumption of each demand component, depending on the model used: the black line corresponds to the basic demand model (assumed demands from billing), and the green line with triangles corresponds to the demand components model.

Table 6.21 contains the times required for the calibration process. It can be seen that the process can be performed in pseudo or quasi real time, as the computational time required for each hour is 20 minutes (51 hours required to calibrate 168 hours).

6.3.7 Discussion

The basic demand model has been used to simulate the network with the precalibration dataset. Figure 6.43 shows that the pressure prediction error when using the basic demand model follows a demand-like pattern, as this demand model is not able to assign a different behavior to each zone of the network. The prediction error can be improved by the use of the demand components model, which

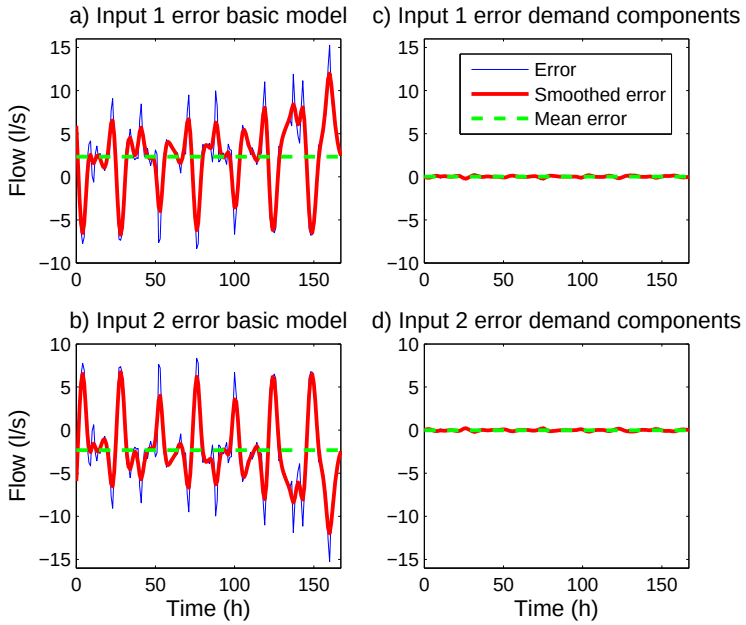

Figure 6.47: Castelldefels Platja flow prediction error at each network input (rows of subfigures) using the basic demand model (first column of subfigures), and the demand components model (second column of subfigures). The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

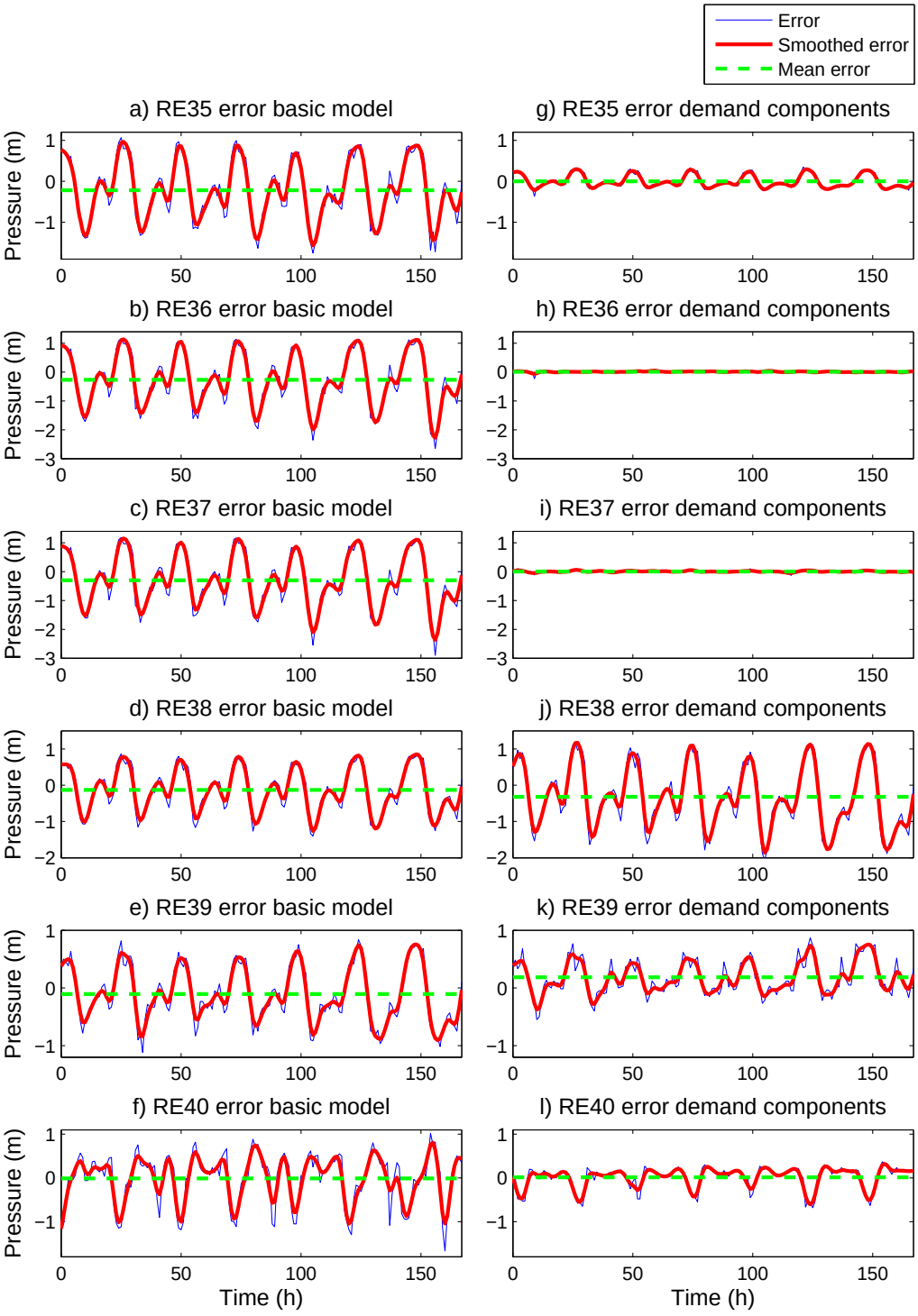

Figure 6.48: Castelldefels Platja pressure prediction error at each sensor (rows of subfigures) using the basic demand model (first column of subfigures), and the demand components model (second column of subfigures). The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

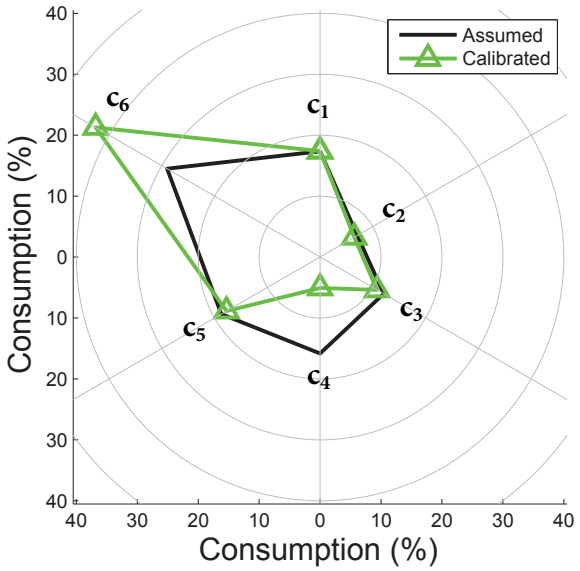

Figure 6.49: Average percentage of water consumption computed from billing (black line), and calibrated demand components (green line with triangles) in Castelldefels Platja network

allows having multiple demand behaviors depending on the location of nodes in the network.

Results in Table 6.19 show that the use of demand components to model nodal demands minimizes the RMSE in the predicted flows and pressures significantly (40% and 70%, respectively). Table 6.20 proves that the calibrated demand components are valid for future data, although as seen in Section 6.1, the online continuous calibration generates an even lower prediction error.

Figure 6.47 and Figure 6.48 show a graphical comparison of the previous results. The flow prediction error is reduced to irrelevant values. The pressure prediction error is considerably minimized in all sensors, except sensor RE38. This sensor is located near one pressure reduction valve (Figure 6.40), making the pressure modification more difficult. A detailed analysis is presented after the discussion section.

The uncertainty of the calibrated demand components can be analyzed through Figure 6.46. During night-time, due to the nonlinear headloss/flow relation (see Equation 2.2), pressure is less sensitive to demand variations. Consequently, the singular values of the sensitivity matrix are smaller, increasing the calibrated parameters variance as seen in Equation 4.7. However, the uncertainty in this network is lower than in the previous case studies, due to the large water consumption during the whole day (minimum night flow of about 40 l/s).

Table 6.22: Engineering time (h), computational time, and number of EPANET simulation runs for the whole methodology

Process	**Engineer time (h)**	**Computational time**	**EPANET runs**
Precalibration process	1.03	0.553 hours	1728 simulations
Online calibration	0.01	0.3 $\frac{hours}{calibrated\ hour}$	300 $\frac{simulations}{calibrated\ hour}$

Figure 6.49 shows a good accuracy in the average percentage of demand components consumption for components $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$, and $\mathbf{c}_5$. Demand component $\mathbf{c}_6$ consumes more water than expected, obtaining it from demand component $\mathbf{c}_4$. These two demand components are situated side by side, what could explain this behavior. Besides, the proximity of sensor R38 (situated in the area predominated by component $\mathbf{c}_6$) to a network input could also explain this inaccuracy. This will be also analyzed in next section.

Finally, Table 6.22 presents the overall time required for the precalibration and online calibration processes. The time required for each of the methodology processes demonstrates the straightforward and fast application of every step involved in the complete calibration process for any water distribution network, and the viability of the online demand calibration, as the calibration process lasts 20 minutes for each calibrated hour. The times in Table 6.15 have not been included in the overall process required time because the sensors were already installed.

6.3.8 RE38 errors analysis

The errors observed (pressure prediction and average percentage of consumption) related with the sensor RE38 have motivated an analysis of what may produce these inaccuracies. Sensor RE38 is situated near a network water input, where pressure is controlled by means of a PRV (see Figure 6.40). The proximity of the sensor to the point where pressure is fixed should make the measurement similar to this pressure and slightly affected by flow variations. However, as seen in Figure 6.50.a, the sensor measurement (green dots) varies its pressure depending on the global water consumption of the network (Figure 6.50.b). This behavior is explained by the effect of head loss due to a long way from the fixed pressure point. On the other hand, both the sensor pressure prediction with the basic (red dashed line) and the calibrated (blue circles) demand models are nearly identical

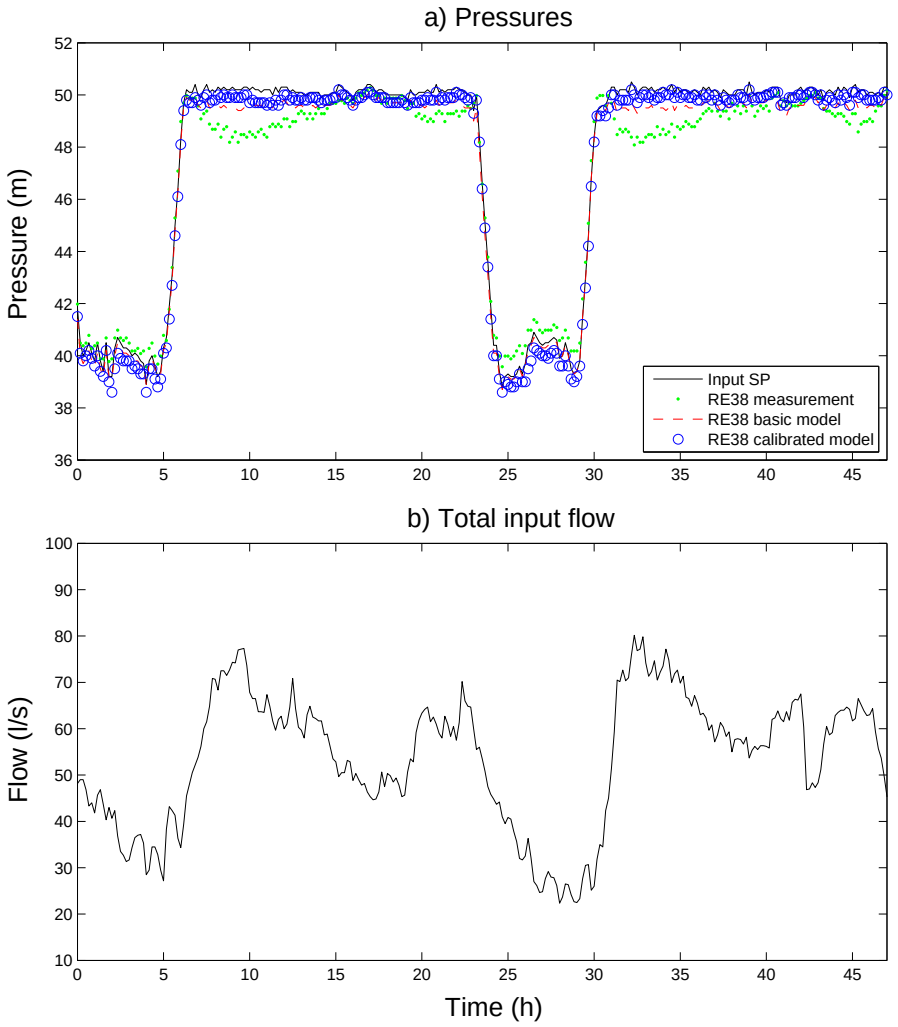

Figure 6.50: Analysis of inaccuracies due to sensor RE38: a) Real and predicted pressures compared to the fixed input pressure set point, and b) Total flow consumed in the network

to the fixed pressure at the PRV (black line).

In conclusion, the inaccuracies in the calibration results suggest a possible closed pipe between the area where sensor RE38 is located and the network input next to that area. The detection of this type of model structural errors is pointed out as an important factor for a good calibration in Walski et al. [2014]. The potential closed pipe is depicted in Figure 6.51 with a red striped line.

The whole calibration process has been repeated considering the pipe indicated in Figure 6.51 as a closed-pipe. The data analysis and depth correction, parameterization and calibration processes have been performed with the same data used when the pipe was considered to be open. Due to the change in the

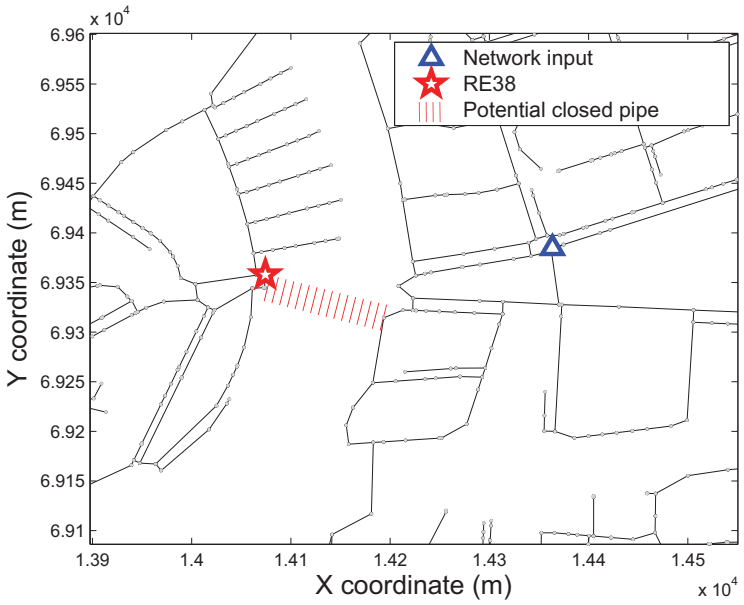

Figure 6.51: Detailed network structure with signaled sensor RE38 (red star), network input (blue triangle) and potential closed pipe (red striped line)

network topology, the zones affected by each demand component vary, and the calibration results are different. Figure 6.52 presents the flow error with the basic demand model and the demand components model. The error is reduced in a similar way as when the pipe was considered to be opened. Figure 6.53 presents the pressure prediction error in the installed sensors. It can be seen that the consideration of the closed pipe leads to a lower prediction error in sensor RE38 compared to the one obtained with the opened pipe (Figure 6.48).

Finally, Figure 6.54 presents the average percentage of demand components' water consumption. The consideration of the closed pipe leads to a better accuracy in the percentage of water consumption.

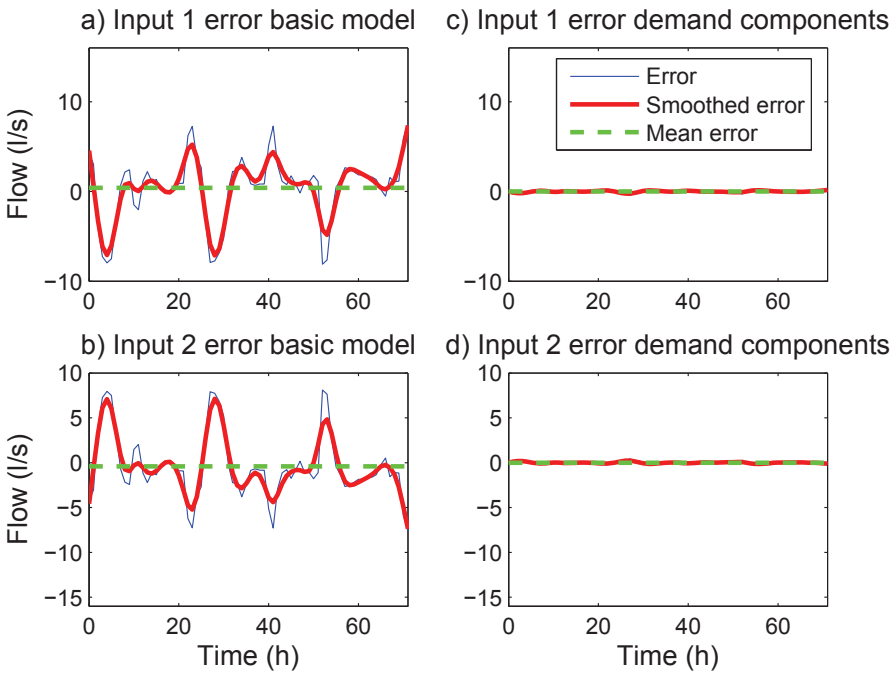

Figure 6.52: Castelldefels Platja (with closed pipe) flow prediction error at each network input (rows of subfigures) using the basic demand model (first column of subfigures), and the demand components model (second column of subfigures). The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

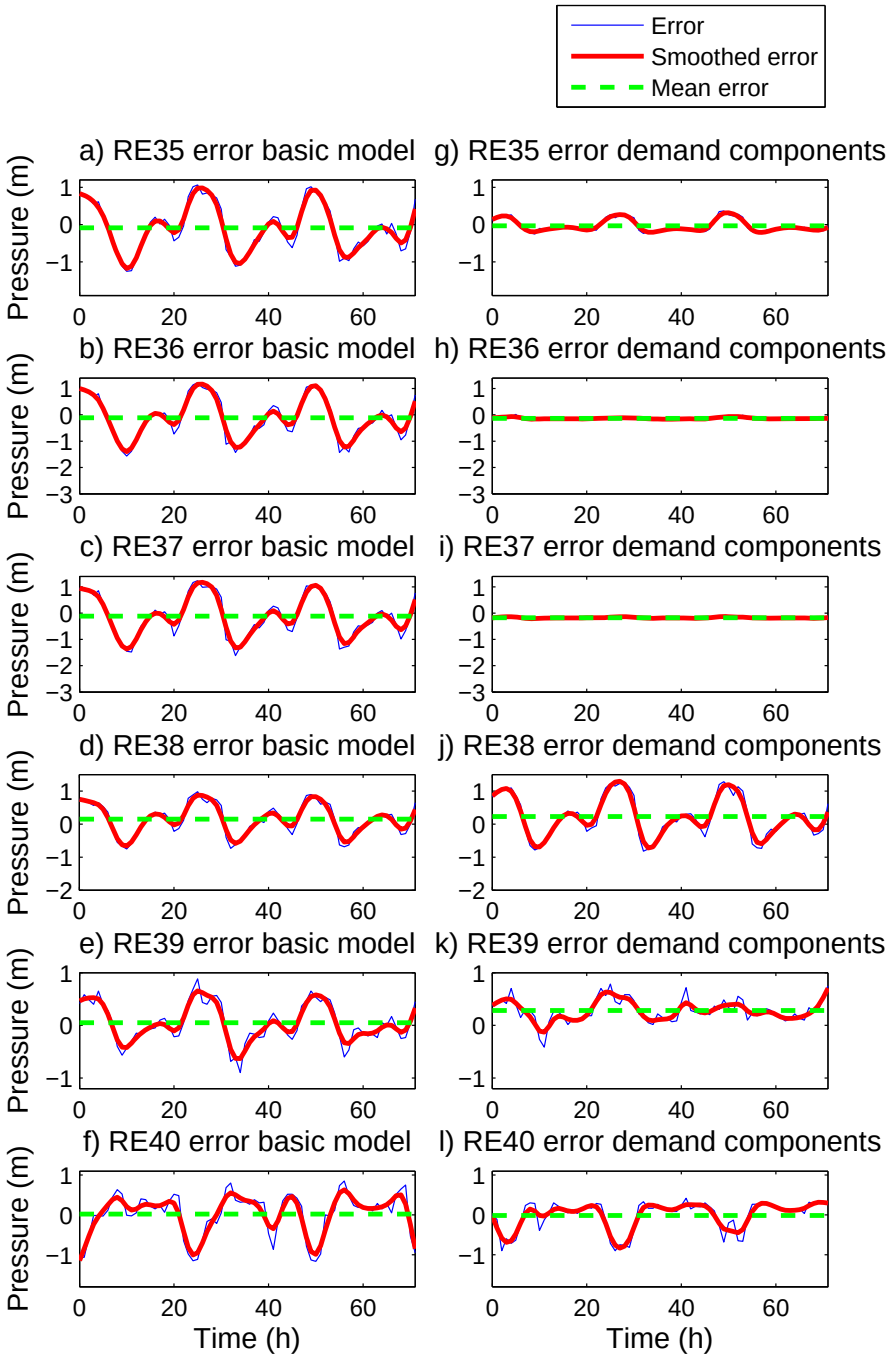

Figure 6.53: Castelldefels Platja (with closed pipe) pressure prediction error at each sensor (rows of subfigures) using the basic demand model (first column of subfigures), and the demand components model (second column of subfigures). The blue thin line corresponds to the raw error, the red thick line corresponds to the smoothed error computed by means of a smoothing spline, and the dashed green line corresponds to the mean error

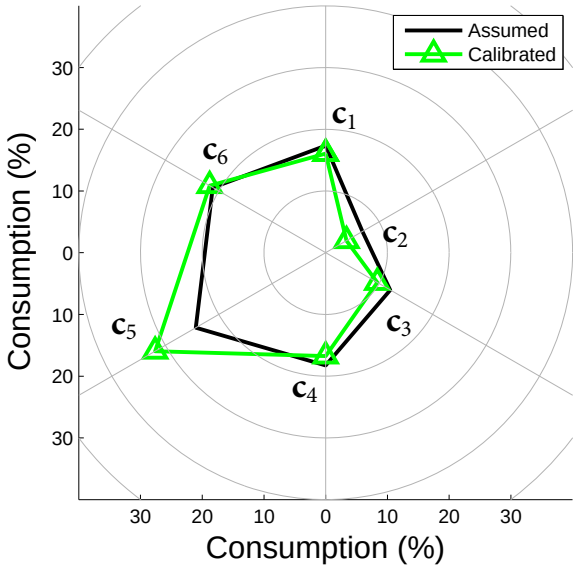

Figure 6.54: Average percentage of water consumption computed from billing (black line), and calibrated demand components (green line with triangles) in Castelldefels Platja network (with closed pipe)

Chapter 7

Concluding remarks

The conclusions and contributions drawn from this thesis are presented in this chapter. First, the main contributions are summarized. In the conclusions section, general conclusions related to the calibration problem are detailed, taking into account the future work derived from the literature review in Chapter 2. Finally, some guidelines are given for future works and extensions.

7.1 Summary of contributions

This thesis has presented the development of an adaptive model that calibrates its parameters online, while discerning between demand evolution and leakage appearance. This model includes some factors that were ignored in previous works focused on calibration: the necessity of the calibration to be performed online to take into account the continuous evolution of the system, and the consideration of the existence of leakage and the possibility of detecting and locating leaks through the calibration process. The main contributions are next summarized:

1. **Definition of a novel demand model based on demand components**: One of the main contributions of this thesis is the demand components model. The main concept of demand components is to represent the demand behavior of areas within the DMA, instead of individual demand behaviors, whose calibration is unattainable.

2. **Use of SVD analysis to define memberships of nodal demands to demand components**: A great improvement with respect to the work done before the thesis is the definition of memberships. The memberships give a measure of pertinence of each node to each demand component,

depending on the sensitivity of this demand to variations on the network variables (flow/head). This permits to generate thousands of different nodal demands from only a few demand components.

3. **Use of SVD analysis to select sensors locations**: The same technique used to generate the memberships can be used to group sensors depending on their sensitivity to the defined demand components, and to select, for each group of sensors, the sensor that has highest sensitivity to one parameter and low sensitivity to the others.

4. **Online application of demand components calibration**: An online calibration structure is proposed. This structure permits to keep an updated model that changes its parameters to adapt to the continuous evolution of the real system. The model can be used for applications that require a well calibrated hydraulic model.

5. **Exploitation of the online demand components calibration for leak detection and localization**: Leakage has to be considered when calibrating the model. The online calibration of geographical demand components is utilized to detect and locate leaks by comparing the calibrated components values with the historical ones, and evaluating if changes are provoked by a system evolution or by a leak.

6. **Standarization of the methodology for straightforward application to WDN**: The development of the whole method has considered that the final application must be able to be applied to any water distribution network.

7. **Programming of a modular software that includes the developed methodologies**: The methodologies developed are being encapsulated in a Matlab GUI for academic purposes.

7.2 Conclusions

The literature review in Chapter 2 conditioned the objectives definition and the working plan included in the thesis proposal. The guidelines extracted from the future work pointed out by Ostfeld et al. [2012] in the summary of the Battle of the Water Calibration Networks (BWCN), are used next to organize the conclusions extracted:

- "The solutions that provide a good match between measured and modeled data have to be validated with extra data": All the case studies in Chapter 6 include the validation of the calibrated parameters with extra data, either from future samples or from sensors not used in the calibration process.
- "Uncertainty has to be included in the model parameters to explore the influence on the calibrated model outputs": The uncertainty propagation has been computed and deeply analyzed to observe the effect of uncertainty in measurements on the uncertainty in parameters under multiple conditions.
- "Calibration size problem reduction is an important factor to consider to avoid model over-fitting, avoid unnecessary simulations or reducing the search space": The parameterization process that represents nodal demands as a combination of demand components is used to reduce the calibration size problem.
- "Leakage data may be included in hydraulic calibration efforts because leakage directly affects nodal demand allocation and pump curve characterizations": The adaptive model is calibrated online, thus leaks are considered part of the calibration. The leak detection and location process is responsible for detecting the variations in demand components due to leakage and for triggering an alarm if this situation happens.
- "The effect of different field data on model calibration should be investigated (use of flow and/or pressure measurements)": Although not implicitly included in this thesis, the author has analyzed the effect of using flow and/or pressure measurements for demand components calibration in Sanz and Pérez [2015b]. Pressure measurements seem to be more reliable for the calibration of geographic demand components.

Apart from considering the particular guidelines given in Ostfeld et al. [2012], three general conclusions have been extracted. First of all, the demand components model proposed changes the point of view about calibrating demands. Instead of calibrating demands so that their individual behavior is as similar as possible to the real behavior, the objective of the demand components model is to calibrate groups of demands that have geographical and hydraulic similarities (the hydraulic similarity is directly related with the network topology, and hence, to the geographical location of the demands). The calibrated demand components generate individual demands that may not be exactly as the real ones, but the aggregated demand in a zone at a specific sample, and the cumulative

demand of each individual node during a period of time (similar to the billing) should coincide with the real ones if other parameters (roughness, valve status, etc.) are well calibrated. Additionally, the demand components model permits to reproduce the uncertainty measured in a real system, as it was presented in Pérez et al. [2015].

Secondly, the analysis of the SVD allows a deep understanding of the system, and the relation between inputs and outputs. These information can be used to parameterize the model, select the sensors' locations, compute the inverse of the sensitivity matrix for the calibration process, and calculate the uncertainty propagation from measurements to parameters, among many other applications that have not been used in this thesis.

Finally, the consideration of the calibration and leak detection as a symbiosis between both processes generates a mutual benefit for them. Leakage is intentionally included in the calibration process, as it is actually a demand that modifies the hydraulic behavior of the system, and is this inclusion which allows to detect and locate leaks thanks to the continuous calibration of geographically distributed demand components.

Focusing on the particular aspects of the work presented, several conclusions can be extracted. These conclusions are presented in the order that would be followed when applying the methodology to a new network.

Initially, the network size is analyzed to evaluate the advantages of reducing the model size by means of the skeletonization process presented in Appendix A. This work includes both a network (Section 6.1) where model reduction is not necessary due to the already low number of junctions and pipes, and two networks (Section 6.2 and Section 6.3), where the huge amount of elements would hinder the viability of the real-time procedure due to the exponential increase of the computational time required. In these two particular case studies, it has been proved that the reduction of the model based on joining pairs of pipes connecting null demand nodes, reduction of extremal nodes connected by a unique pipe, and reduction of parallel pipes, does not affect the hydraulic behavior of the model.

After the network model reduction, the precalibration process consisting of the definition of parameters and the selection of sensors locations is performed. Both processes require the computation of the sensitivity matrix. This thesis uses the method presented in Section 2.2. This approach is similar to the sensitivity equation method, but including a linearization process that allows to reduce the number of simulation runs required.

Then, the parameterization process defines the memberships of each nodal demand (initial parameter) to each demand component (new parameter) by means of the analysis of the SVD of the sensitivity matrix. This analysis generates groups of nodal demands that have similar sensitivities, defining the membership values depending on these similarities. The memberships can be binary values, positive values, or values with no restriction. Positive values give the best result in terms of calibration prediction error (Sanz and Pérez [2015a]).

The same process is used to define the sensors needed to obtain a determined system of equations. The methodology proposed is able to consider flow/pressure sensors. The analysis of the sensitivity matrix SVD finds the highest sensitive sensor among groups of sensors that have similar sensitivities. The study in Sanz and Pérez [2015b] compares the A, D, and V-optimalities from the proposed methodology solution to a significant set of possible solutions, concluding that results can be little improved when considering pressure sensors or a combination of flow and pressure sensors.

Results in Sanz and Pérez [2015b] show that measuring pressure seems to be the best option when calibrating demand components, whereas metering flow requires a higher number of sensors to achieve a good calibration. This can be justified by the meshed topology of the network: pressure is more representative of a geographical zone, thus improving predicted pressure in a particular point of the network will improve predicted pressure in the nearby locations; on the other hand, improving the predicted flow at a particular pipe does not necessarily improves the rest of flows of the zone, due to the meshed topology of the network.

The precalibration process generates all the required information for the online calibration of demand components: sensors from where the network data will be extracted, and the memberships of nodal demands to demand components that reduce the parameters to be calibrated. The online calibration process computes then the demand components values that minimize the measurements prediction errors for a determined number of samples.

The calibrated demand components model reduces the prediction error for all measurements in the three case studies presented, compared to the basic demand model, which generates a pressure prediction error that follows the total water consumption profile. This type of error is provoked by the constant distribution of water among all demands, which is overcome by the demand components model.

The author proposes a gradient-type implicit method that solves the generalized inverse problem using the SVD to compute the inverse of the sensitivity matrix. Results are compared with the solution from the GAs. The gradient-type

approach generates better results in terms of prediction error minimization, validity of calibrated parameters for future samples, and computational time required. One conclusion extracted is that a high degree of knowledge and experience is required to set the parameters of the GAs in order to obtain good results, whereas the gradient-type methods do not require these high degree of skill.

The uncertainty propagation from sensors' measurements to calibrated parameters values is computed by means of the FOSM model. Results obtained in Section 6.1 show that the use of data from multiple days to calibrate a single set of demand components values reduce the uncertainty in the calibrated values particularly during nighttime, when the noise in measurements has a greater impact due to the low water consumption.

Related to the number of samples used (calibration horizon), results from Section 6.1 show that the higher the value of the calibration horizon, the higher the validity of the demand components values for future data. However, the prediction error for current data becomes higher, as daily events are not completely captured, i.e. their effect is somewhat reduced by the other samples used. The decision on the setting of the calibration horizon depends on the final use of the model: large values are appropriate for models that will be used for prediction of future network state, whereas small values are suitable for models that want to capture daily changes and particular events that occur in the network (e.g. leakage).

Finally, the leak detection and location method based on demand components calibration (Chapter 5) is applied to a real scenario in subsection 6.2.4. The analysis of the calibrated demand components allow to detect the leak only four hours after its appearance, with an accuracy of 167 meters. Both proposed localization approaches (direct and leak membership) generate similar localization accuracies. These results improve the ones from Pérez et al. [2014], where the leak was located with similar accuracy after 10 hours from its appearance. The leak detection and localization methodology has been further analyzed in subsection 6.2.3 using synthetic data. The main conclusion extracted is that leaks whose effect is lower than the demand components uncertainty cannot be detected. A potential solution is to reduce the size of demand components by introducing more sensors that allow the definition of a higher number of demand components that cover smaller network areas.

The modular and step-based approach proposed gives the versatility to the user of replacing any of the methodologies used for his/her own developed approaches. The sensitivity matrix computation, the sensor placement method-

ology, the optimization method used, the uncertainty model, the indicators for leak detection, etc., can be redefined to fulfill particular user necessities. The developed methodology has been encapsulated in the existing software. The last version of the software is available in Sanz and Pérez [2013a]. This version only includes the demand calibration module.

To conclude, the developed methodologies can be directly applied to other pressurized water networks, such as irrigation systems, where only few modifications would be needed. The methodologies could also be used, with a higher amount of dedication, in gas distribution networks or similar complex networks.

7.3 Future work

As noted in the review of the state of the art (Chapter 2), research in WDN calibration is continuously advancing, and no unique and best methodology has been found. The proposed method bases on a redefinition of the demand model, which fulfills most of the possible uses of water distribution network hydraulic models. However, a lot of work can still be done to improve the proposed methodology.

The author has performed an experimental analysis of the A-, D-, and V-optimalities of the sensor placement methodology. An analytic analysis would completely validate the proposed approach over other sampling design methods.

Referring to the optimization method used, the gradient-type solution has generated, for the presented case studies, better results than the GAs. However, the GAs have proved to be one of the best optimization methods, thus the author presumes that the results obtained with this method can be improved with a better GAs parameter setting. Further work has to be done to implement the proposed demand components model with the GAs optimization. On the other hand, other non-evolutionary methods have to be tested: the Gauss-Newton method may improve the results obtained because it consists in making a scaling with second order information, or the Levenberg-Marquardt method, which consists in making a damping on the diagonal of the iteration matrix to ensure step size control and gradually turn the descent direction to the gradient one.

The leak detection method has proved to correctly detect and locate a real leak. Furthermore, the analysis with synthetic data has extracted additional conclusions about detectability of leaks. However, the method has to be further tested. Scenarios including multiple leaks can be analyzed to determine the abil-

ity to detect simultaneous burst. Additionally, flow sensors can be tested and compared with pressure sensors in order to assess which is the best option.

The leak detection method has to be tested to check the ability to detect not only burst, but unknown status valves, model structural errors, and other events that affect the hydraulic behavior of the network.

Finally, and as a near future work, the author's objective is to complete the software with the whole methodology, making it available in the website for educational purposes.

Appendix A

Network skeletonization

A.1 Motivation

Network skeletonization consists in selecting for inclusion in the model only the parts of the hydraulic network that have a significant impact on the behavior of the system (Walski et al. [2003]). The level of skeletonization depends on the intended use of the model. Network skeletonization is proposed in this thesis for two main reasons:

- Reduction of the computational time
- Elimination of mathematical incorrectness

For the first item, it is mandatory for the methodology to be executed in real-time. The number of elements in the WDN model is directly connected with the size of the matrices involved in the calibration problem, as well as the model inputs to be fixed before executing one simulation run. The reduction of the network model elements directly impacts in the computational time required for each of these processes.

On the other hand, network skeletonization helps to avoid mathematical problems in the computation of sensitivity matrices. In Chapter 4, a calibration methodology based on least squares is presented. This methodology uses the nonlinear matrix C whose parameters are calculated as shown in Equation 2.8. When the flow through a pipe is null, the head loss on this pipe is zero.

$$\mathbf{h}_i - \mathbf{h}_j = R_{ij} \cdot \mathbf{q}_{ij} \cdot |\mathbf{q}_{ij}|^{0.852} = R_{ij} \cdot 0 \cdot |0|^{0.852} = 0 \tag{A.1}$$

However, in subsection 2.1.1 (and in the used methodology) matrix $\mathbf{C}$ is used to calculate the flow from the head loss, turning the previous equation into Equa-

tion 2.8, where if the head loss is null:

$$C_{pp} = G_{ij} \cdot |\mathbf{h}_i - \mathbf{h}_j|^{-0.46} = G_{ij} \cdot |0|^{-0.46} = \infty \tag{A.2}$$

Consequently, an infinite value appears in matrix $\mathbf{C}$, making the calculation of $\mathbf{A}^{-1}$ not possible. An intuitive solution is to replace this infinite value for a very high one. However, the replacement of the infinite values may cause some inaccuracies in the results. Going in depth into the problem, the topology of the network seems to be the main cause:

- Nodes connected with very short pipes that have a very low resistivity factor may generate null head losses.
- Series pipes which connect a null demand node.

The application of network skeletonization techniques eliminates these problems while keeping the same (or very similar) hydraulic behavior of the network model.

A.2 Proposed solution

As seen in Chapter 2, there exist many network reduction techniques. In this work, skeletonization based on the steps listed in Walski et al. [2003] is used. First, a basic skeletonization process which does not affect the hydraulic behavior is performed:

1. Pairs of pipes connecting a null demand node are joint.
2. Extremal nodes connected by a unique pipe to the network are reduced.
3. Parallel pipes are replaced by a single equivalent pipe.

This reduction does not affect the hydraulic behavior of the network. The accuracy in results when locating leakages may be slightly affected due to the suppression of intermediate nodes.

The second step consists in the reduction of short pipes with low resistivity factor, connected in series with another pipe. This modification may affect the hydraulic behavior, depending on the intended use of the model. In this process, pipes with a resistivity factor lower than a specified threshold are absorbed by the adjoin pipe, and the junction connecting them eliminated. The consumption

of the eliminated junction is added to the nearest junction. The parameters of the pipe connected to the deleted pipe are recalculated, as seen in Equation A.3:

$$C_r = (\frac{L_r}{D_r^{4.87}})^{0.54}(\sum_{i=1}^{2} \frac{L_i}{D_i^{4.87} C_i^{1.852}})^{-0.54} \tag{A.3}$$

where C_r is the roughness of the new pipe; L_r is the length of the new pipe, equal to $L_1 + L_2$; D_r is the diameter of the new pipe; and L_i, D_i and C_i are the length, diameter and roughness of pipe i.

The process is performed iteratively, calculating the resistivity of each pipe of the network, assessing if the value is beyond a defined threshold and, if necessary, eliminating the pipe and reconnecting and reparametrizing the appropriate elements of the network.

Appendix B

Resolution matrix and information density matrix

The quality of the generalized inverse solution can be characterized by means of the resolution matrix. Starting from any matrix $\mathbf{S}$ relating demands $\mathbf{d}$ with heads $\mathbf{h}$, the latter can be calculated by solving $\mathbf{S} \cdot \mathbf{d}$. Repeating the process backwards, we can multiply $\mathbf{S^{-1}}$ by $\mathbf{h}$ to get back a demand solution $\hat{\mathbf{d}}$:

$$\hat{\mathbf{d}} = \mathbf{S^{-1}S} \cdot \mathbf{d} \tag{B.1}$$

Ideally, $\hat{\mathbf{d}} = \mathbf{d}$. However, since the original model may have had a nonzero projection onto the model null space $N(\mathbf{S})$, $\hat{\mathbf{d}}$ will not in general be equal to $\mathbf{d}$. Then, the model space resolution matrix is given by:

$$\mathbf{R} = \mathbf{S^{-1}S} \tag{B.2}$$

Or in SVD terms

$$\mathbf{R} = \mathbf{V\Lambda}^{-1}\mathbf{U}^T\mathbf{U\Lambda V}^T$$

$$\mathbf{R} = \mathbf{VV^T} \tag{B.3}$$

As explained in Aster et al. [2005], if $N(\mathbf{S})$ is trivial, then $\mathbf{S}$ is full rank and $\mathbf{R}$ is the identity matrix $\mathbf{I_n}$. In this case the original model is recovered exactly and the resolution is perfect. If $N(\mathbf{S})$ is a nontrivial subspace of $\mathbb{R}^n$, then $rank(\mathbf{S}) < n$, so that $\mathbf{R}$ is not the identity matrix. The parameter space resolution matrix is instead a symmetric matrix describing how the generalized inverse solution smears out the original model $\mathbf{d}$, into a recovered model $\hat{\mathbf{d}}$.

The same process can be performed in order to obtain the data space resolu-

tion matrix, also called information density matrix $\mathbf{I_d}$. Starting from an inverse matrix $\mathbf{S}^{-1}$ and the data vector $\mathbf{h}$, the model $\hat{\mathbf{d}}$ can be calculated by solving $\mathbf{S}^{-1} \cdot \mathbf{h}$. Then, the data $\hat{\mathbf{h}}$ can be recovered by multiplying $\mathbf{S} \cdot \hat{\mathbf{d}}$:

$$\hat{\mathbf{h}} = \mathbf{S}\mathbf{S}^{-1} \cdot \mathbf{h} \tag{B.4}$$

Hence, the information density matrix is:

$$\mathbf{I_d} = \mathbf{S}\mathbf{S}^{-1} \tag{B.5}$$

Or in SVD terms

$$\mathbf{I_d} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T$$

$$\mathbf{I_d} = \mathbf{U}\mathbf{U}^T \tag{B.6}$$

If $N(\mathbf{S}^{-1})$ contains only the zero vector, then $rank(\mathbf{S}) = \mathbf{m}$, and $\mathbf{I_d} = \mathbf{I_m}$. In this case, $\hat{\mathbf{h}} = \mathbf{h}$, and the generalized inverse solution $\hat{\mathbf{d}}$ fits the data exactly. However, if $N(\mathbf{S}^{-1})$ is nontrivial, then $rank(\mathbf{S}) < m$, and $\mathbf{I_d}$ is not the identity matrix. In this case, $\hat{\mathbf{h}}$ is not equal to $\mathbf{h}$, and the model $\hat{\mathbf{d}}$ does not exactly fit the data $\mathbf{h}$.

Bibliography

A. D. o. E. C. ADEC. Technical Review of Leak Detection Technologies - vol.1 - Crude Oil Transmission Pipelines. Technical report, Alaska Department of Environmental Conservation, Anchorage, 1999. 2.7

I. Ahmed, K. Lansey, and J. Araujo. Data collection for water distribution network calibration. In *2nd International Conference on Water Pipeline Systems*, pages 271–278, Exeter, 1999. 2.3, 2.5, 2.6

K. Aksela and M. Aksela. Demand Estimation with Automated Meter Reading in a Distribution Network. *Journal of Water Resources Planning and Management*, 137(5):456–467, sep 2011. doi: 10.1061/(ASCE)WR.1943-5452.0000131. 2.3.1

American Water Works Association Research Comittee on Distribution Systems. Water Distribution Research and Applied Development Needs. *Journal of the American Water Works Association*, 66(6):385–390, 1974. 2.3

E. J. Anderson and K. H. Al-Jamal. Hydraulic-Network Simplification. *Journal of Water Resources Planning and Management*, 121(3):235–240, may 1995. doi: 10.1061/(ASCE)0733-9496(1995)121:3(235). 2.1.4

F. Arbués, M. Á. Garca-Valiñas, and R. Martnez-Espiñeira. Estimation of residential water demand: a state-of-the-art review. *The Journal of Socio-Economics*, 32(1):81–102, mar 2003. doi: 10.1016/S1053-5357(03)00005-2. 2.3.1

R. Aster, B. Borchers, and C. Thurber. *Parameter Estimation and Inverse Problems*. Elsevier, New York, 2005. 3.1, 3.3, 3.4, 4.1, 4.5, B

Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, San Diego, California, 1974. 2.2, 2.5, 4.1

A. Bargiela. An algorithm for observability determination in water-system state estimation. *Control Theory and Applications*, 132(6):245–250, 1985. 2.4

A. Bargiela and G. Hainsworth. Pressure and Flow Uncertainty in Water Systems. *Journal of Water Resources Planning and Management*, 115(2):212–229, mar 1989. doi: 10.1061/(ASCE)0733-9496(1989)115:2(212). 2.5

K. Behzadian, Z. Kapelan, D. Savic, and A. Ardeshir. Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks. *Environmental Modelling & Software*, 24(4):530–541, apr 2009. doi: 10.1016/j.envsoft.2008.09.013. 2.5, 2.6

Bentley. WaterCAD Web Page. URL `http://www.bentley.com/en-US/Products/WaterCAD/`. 2.1.3

P. Bhave. Calibrating Water Distribution Network Models. *Journal of Environmental Engineering*, 114(1):120–136, feb 1988. doi: 10.1061/(ASCE)0733-9372(1988)114:1(120). 2.3.2

E. Bonada, Jordi Meseguer, and J. M. M. Tur. Practical-Oriented Pressure Sensor Placement for Model-Based Leakage Location in Water Distribution Networks. In Michael Piasecki, editor, *Informatics and the Environment: Data and Model Integration in a Heterogeneous Hydro World*, New York, 2014. 6.1.1, 6.3.3

P. F. Boulos and D. J. Wood. Explicit Calculation of Pipe Network Parameters. *Journal of Hydraulic Engineering*, 116(11):1329–1344, nov 1990. doi: 10.1061/(ASCE)0733-9429(1990)116:11(1329). 2.3.2

T. Brandon. *Water Distribution Systems (Water Practice Manuals)*. The Insitute of Water Engineers and Scientists, London, 1984. 2.1.4

M. A. Brdys and B. Ulanicki. *Operational Control of Water Systems: Structures, Algorithms and Applications*. Prentice Hall International, 1994. 2.1, 2.1.1, 2.1.1, 2.1.3

D. Broad, H. Maier, and G. Dandy. Optimal Operation of Complex Water Distribution Systems Using Metamodels. *Journal of Water Resources Planning and Management*, 136(4):433–443, jul 2010. doi: 10.1061/(ASCE)WR.1943-5452.0000052. 2.1.4

C. Bush and J. Uber. Sampling Design Methods for Water Distribution Model Calibration. *Journal of Water Resources Planning and Management*, 124(6): 334–344, nov 1998. doi: 10.1061/(ASCE)0733-9496(1998)124:6(334). 2.5, 2.6

P. Carpentier and G. Cohen. State estimation and leak detection in water distribution networks. *Civil Engineering Systems*, 8(4):247–257, dec 1991. doi: 10.1080/02630259108970634. 2.4

P. Carpentier, G. Cohen, and Y. Hamam. *Water Network Equilibrium, Variational Formulation and Comparison of Numerical Algorithms*, volume 1. New York, wiley edition, 1987. 2.1.3

A. Cesario and T. Lee. A Computer Method for Loading Model Networks. *Journal of the American Water Works Association*, 72(4):208–213, 1980. 2.3.1

L. C. Chen. *Pipe network transient analysis - The forward and inverse problems.* PhD thesis, Cornell University, 1995. 2.2

W. Cheng and Z. He. Calibration of Nodal Demand in Water Distribution Systems. *Journal of Water Resources Planning and Management*, 137(1):31–40, jan 2011. doi: 10.1061/(ASCE)WR.1943-5452.0000093. 2.2, 2.2.1, 2.3.1, 2.3.2, 4.1, 4.4

A. F. Colombo, P. Lee, and B. W. Karney. A selective literature review of transient-based leak detection methods. *Journal of Hydro-environment Research*, 2(4):212–227, apr 2009. doi: 10.1016/j.jher.2009.02.003. 2.7

E. Creaco, R. Farmani, Z. Kapelan, L. Vamvakeridou-Lyroudia, and D. Savic. Considering the Mutual Dependence of Pulse Duration and Intensity in Models for Generating Residential Water Demand. *Journal of Water Resources Planning and Management*, 141(11):04015031, nov 2015. doi: 10.1061/(ASCE)WR.1943-5452.0000557. 2.3.1

R. Datta and K. Sridharan. Parameter Estimation in Water Distribution Systems by Least Squares. *Journal of Water Resources Planning and Management*, 120 (4):405–422, jul 1994. doi: 10.1061/(ASCE)0733-9496(1994)120:4(405). 2.3.2, 2.5

J. Davidson and F. Bouchart. Adjusting Nodal Demands in SCADA Constrained Real-Time Water Distribution Network Models. *Journal of Hydraulic Engineering*, 132(1):102–110, jan 2006. doi: 10.1061/(ASCE)0733-9429(2006)132:1(102). 2.3.1, 2.3.2

W. de Schaetzen, G. Walters, and D. Savic. Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms. *Urban Water Journal*, 2(2):141–152, jun 2000. doi: 10.1016/S1462-0758(00)00052-2. 2.6

G. Del Giudice and C. Di Cristo. Sampling design for water distribution networks. *Transactions on Ecology and the Environment*, 61, 2003. doi: 10.2495/WRM030181. 2.6

J. Deuerlein. Decomposition Model of a General Water Supply Network Graph. *Journal of Hydraulic Engineering*, 134(6):822–832, jun 2008. doi: 10.1061/(ASCE)0733-9429(2008)134:6(822). 2.1.4

C. Eggener and L. Polwoski. Network Models and the Impact of Modeling Assumptions. *Journal of the American Water Works Association*, 68(4):189–196, 1976. 2.1.4, 2.3

EPA. EPANET Web Page. URL `http://www.epa.gov/nrmrl/wswrd/dw/epanet.html`. 2.1.3

B. Farley, S. R. Mounce, and J. B. Boxall. Field Validation of "Optimal" Instrumentation Methodology for Burst/Leak Detection and Location. In *Water Distribution Systems Analysis 2010*, pages 1093–1102, Reston, VA, dec 2011. American Society of Civil Engineers. doi: 10.1061/41203(425)99. 2.7

G. Ferreri, E. Napoli, and A. Tumbiolo. Calibration of roughness in water distribution systems. In *2nd International Conference on Water Pipeline Systems*, pages 379–396, 1994. 2.3.2, 2.6

O. Giustolisi and T. Walski. Demand Components in Water Distribution Network Analysis. *Journal of Water Resources Planning and Management*, 138(4):356–367, jul 2012. doi: 10.1061/(ASCE)WR.1943-5452.0000187. 3.2

D. Goldberg. *Genetic algorithms in search, optimization and machine learning.* Reading, 1989. 2.3.2

J.-A. Goulet, S. Coutu, and I. F. Smith. Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. *Advanced Engineering Informatics*, 27(2):261–269, apr 2013. doi: 10.1016/j.aei.2013.01.001. 2.3, 2.7, 4.6

M. Greco and G. Del Giudice. New Approach to Water Distribution Network Calibration. *Journal of Hydraulic Engineering*, 125(8):849–854, aug 1999. doi: 10.1061/(ASCE)0733-9429(1999)125:8(849). 2.3.2

A. Griewank, D. Juedes, H. Mitev, J. Utke, O. Vogel, and A. Walther. ADOL-C: A package for the automatic differentiation of algorithms written in C/C++, 1998. 2.2

D. Hamberg and U. Shamir. Schematic Models for Distribution Systems Design. I: Combination Concept. *Journal of Water Resources Planning and Management*, 114(2):129–140, mar 1988. doi: 10.1061/(ASCE)0733-9496(1988)114:2(129). 2.1.4

J.-M. Henault, G. Moreau, S. Blairon, J. Salin, J.-R. Courivaud, F. Taillade, E. Merliot, J.-P. Dubois, J. Bertrand, S. Buschaert, S. Mayer, and S. Delepine-Lesoille. Truly Distributed Optical Fiber Sensors for Structural Health Monitoring: From the Telecommunication Optical Fiber Drawling Tower to Water Leakage Detection in Dikes and Concrete Structure Strain Monitoring. *Advances in Civil Engineering*, 2010:1–13, 2010. doi: 10.1155/2010/930796. 2.7

J. Hugenschmidt and A. Kalogeropoulos. The inspection of retaining walls using GPR. *Journal of Applied Geophysics*, 67(4):335–344, apr 2009. doi: 10.1016/j.jappgeo.2008.09.001. 2.7

C. J. Hutton, Z. Kapelan, L. Vamvakeridou-Lyroudia, and D. A. Savić. Dealing with Uncertainty in Water Distribution System Models: A Framework for Real-Time Modeling and Data Assimilation. *Journal of Water Resources Planning and Management*, 140(2):169–183, feb 2014. doi: 10.1061/(ASCE)WR.1943-5452.0000325. 2.3, 2.5

D. D. Jackson. Interpretation of Inaccurate, Insufficient and Inconsistent Data. *Geogphysical Journal of the Royal Astronomical Society*, 28(2):97–109, 1972. doi: 10.1111/j.1365-246X.1972.tb06115.x. 4.1, 4.1

D. Kang and K. Lansey. Real-Time Demand Estimation and Confidence Limit Analysis for Water Distribution Systems. *Journal of Hydraulic Engineering*, 135(10):825–837, oct 2009. doi: 10.1061/(ASCE)HY.1943-7900.0000086. 2.3.2, 2.5

D. Kang and K. Lansey. Optimal Meter Placement for Water Distribution System State Estimation. *Journal of Water Resources Planning and Management*, 136 (3):337–347, may 2010. doi: 10.1061/(ASCE)WR.1943-5452.0000037. 2.6

D. Kang and K. Lansey. Demand and Roughness Estimation in Water Distribution Systems. *Journal of Water Resources Planning and Management*, 137(1): 20–30, jan 2011. doi: 10.1061/(ASCE)WR.1943-5452.0000086. 2.3.2

Z. Kapelan. *Calibration of water distribution system hydraulic models.* PhD thesis, University of Exeter, 2002. 2.3.2

Z. Kapelan, D. Savic, and G. Walters. A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks. *Journal of Hydraulic Research*, 41(5):481–492, sep 2003a. doi: 10.1080/00221680309499993. 2.1, 2.6

Z. Kapelan, D. Savic, and G. Walters. Multiobjective Sampling Design for Water Distribution Model Calibration. *Journal of Water Resources Planning and Management*, 129(6):466–479, nov 2003b. doi: 10.1061/(ASCE) 0733-9496(2003)129:6(466). 2.2, 2.5, 2.5.1, 2.6

Z. Kapelan, D. Savic, G. Walters, D. Covas, I. Graham, and C. Maksimovic. An assessment of the application of inverse transient analysis for leak detection: Part I. In C. Maksimovic, D. Butler, and Memon, editors, *International Conference on Advances in Water Supply Management*, Lisse, 2003c. Swets and Zeitlinger. 2.7

Z. Kapelan, D. Savic, and G. Walters. Optimal Sampling Design Methodologies for Water Distribution Model Calibration. *Journal of Hydraulic Engineering*, 131(3):190–200, mar 2005. doi: 10.1061/(ASCE)0733-9429(2005)131:3(190). 2.5, 2.6

S. H. Kim. Extensive Development of Leak Detection Algorithm by Impulse Response Method. *Journal of Hydraulic Engineering*, 131(3):201–208, mar 2005. doi: 10.1061/(ASCE)0733-9429(2005)131:3(201). 2.7

G. Krumpholz, K. Clements, and P. Davis. Power System Observability: A Practical Algorithm Using Network Topology. *IEEE Transactions on Power Apparatus and Systems*, PAS-99(4):1534–1542, jul 1980. doi: 10.1109/TPAS. 1980.319578. 2.4

A. Lambert. Accounting for Losses: The Bursts and Background Concept. *Water and Environment Journal*, 8(2):205–214, apr 1994. doi: 10.1111/j.1747-6593.1994.tb00913.x. 2.3.1

K. Lansey and C. Basnet. Parameter Estimation for Water Distribution Networks. *Journal of Water Resources Planning and Management*, 117(1):126, 1991. doi: 10.1061/(ASCE)0733-9496(1991)117:1(126). 2.3.2

K. Lansey, W. El-Shorbagy, I. Ahmed, J. Araujo, and C. Haan. Calibration Assessment and Data Collection for Water Distribution Networks. *Journal of Hydraulic Engineering*, 127(4):270–279, apr 2001. doi: 10.1061/(ASCE)0733-9429(2001)127:4(270). 2.3.2, 2.5, 2.6

J. Liggett and L. Chen. Inverse Transient Analysis in Pipe Networks. *Journal of Hydraulic Engineering*, 120(8):934–955, aug 1994. doi: 10.1061/(ASCE)0733-9429(1994)120:8(934). 2.1, 2.3.2, 2.6, 2.7

S. Lingireddy and L. E. Ormsbee. Optimal Network Calibration Model Based on Genetic Algorithms. In *WRPMD'99*, pages 1–8, Reston, VA, jun 1999. American Society of Civil Engineers. doi: 10.1061/40430(1999)45. 2.3.2

H. Loaiciga, R. Charbeneau, L. Everett, G. Fogg, B. Hobbs, and S. Rouhani. Review of Ground Water Quality Monitoring Network Design. *Journal of Hydraulic Engineering*, 118(1):11–37, jan 1992. doi: 10.1061/(ASCE)0733-9429(1992)118:1(11). 2.6

H. Maier, Z. Kapelan, J. Kasprzyk, J. Kollat, L. Matott, M. Cunha, G. Dandy, M. Gibbs, E. Keedwell, A. Marchi, A. Ostfeld, D. Savic, D. Solomatine, J. Vrugt, A. Zecchin, B. Minsker, E. Barbour, G. Kuczera, F. Pasha, A. Castelletti, M. Giuliani, and P. Reed. Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. *Environmental Modelling & Software*, 62:271–299, dec 2014. doi: 10.1016/j.envsoft.2014.09.013. 2.3.2

K. Mallick, I. Ahmed, K. Tickle, and K. Lansey. Determining Pipe Groupings for Water Distribution Networks. *Journal of Water Resources Planning and Management*, 128(2):130–139, mar 2002. doi: 10.1061/(ASCE)0733-9496(2002)128:2(130). 2.5

F. Martinez, B. Ulanicki, and E. Salomons. Fast and Practical Method for Model Reduction of Large-Scale Water-Distribution Networks. *Journal of*

Water Resources Planning and Management, 140(4):444–456, 2014. doi: 10.1061/(ASCE)WR.1943-5452.0000333. 2.1.4

R. Meier and B. Barkdoll. Sampling Design for Network Model Calibration Using Genetic Algorithms. *Journal of Water Resources Planning and Management*, 126(4):245–250, jul 2000. doi: 10.1061/(ASCE)0733-9496(2000)126:4(245). 2.6

W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, 1982. 4.1

S. R. Mounce, J. B. Boxall, and J. Machell. Development and Verification of an Online Artificial Intelligence System for Detection of Bursts and Other Abnormal Flows. *Journal of Water Resources Planning and Management*, 136(3): 309–318, may 2010. doi: 10.1061/(ASCE)WR.1943-5452.0000030. 2.7

S. R. Mounce, R. B. Mounce, and J. B. Boxall. Novelty detection for time series data analysis in water distribution systems using support vector machines. *Journal of Hydroinformatics*, 13(4):672, oct 2011. doi: 10.2166/hydro.2010.144. 2.7

F. Nejjari, R. Pérez, V. Puig, J. Quevedo, M. Cugueró, G. Sanz, and J. Mirats. Abnormal quality detection and isolation in water distribution networks using simulation models. In *11th International Conference on Computing and Control for the Water Industry*, pages 461–466, Exeter, 2011. 4.6

F. Nejjari, R. Sarrate, and J. Blesa. Optimal Pressure Sensor Placement in Water Distribution Networks Minimizing Leak Location Uncertainty. *Procedia Engineering*, 119:953–962, 2015. doi: 10.1016/j.proeng.2015.08.979. 2.6

L. Ormsbee. Implicit Network Calibration. *Journal of Water Resources Planning and Management*, 115(2):243–257, mar 1989. doi: 10.1061/(ASCE) 0733-9496(1989)115:2(243). 2.3, 2.3.2

L. Ormsbee and D. Wood. Explicit Pipe Network Calibration. *Journal of Water Resources Planning and Management*, 112(2):166–182, mar 1986. doi: 10.1061/(ASCE)0733-9496(1986)112:2(166). 2.3.2

A. Ostfeld, E. Salomons, L. Ormsbee, J. Uber, C. Bros, P. Kalungi, R. Burd, B. Zazula-Coetzee, T. Belrain, D. Kang, K. Lansey, H. Shen, E. McBean, Z. Yi Wu, T. Walski, S. Alvisi, M. Franchini, J. Johnson, S. Ghimire, B. Barkdoll, T. Koppel, A. Vassiljev, J. H. Kim, G. Chung, D. G. Yoo, K. Diao, Y. Zhou,

J. Li, Z. Liu, K. Chang, J. Gao, S. Qu, Y. Yuan, T. D. Prasad, D. Laucelli, L. Vamvakeridou Lyroudia, Z. Kapelan, D. Savic, L. Berardi, G. Barbaro, O. Giustolisi, M. Asadzadeh, B. Tolson, and R. McKillop. Battle of the Water Calibration Networks. *Journal of Water Resources Planning and Management*, 138(5):523–532, sep 2012. doi: 10.1061/(ASCE)WR.1943-5452.0000191. 2.3, 7.2

T. Ozawa. The principal partition of a pair of graphs and its applications. *Discrete Applied Mathematics*, 17(1-2):163–186, may 1987. doi: 10.1016/0166-218X(87)90011-4. 2.4

C. V. Palau, F. J. Arregui, and M. Carlos. Burst Detection in Water Networks Using Principal Component Analysis. *Journal of Water Resources Planning and Management*, 138(1):47–54, jan 2012. doi: 10.1061/(ASCE)WR.1943-5452.0000147. 2.7

D. Paluszczyszyn, P. Skworcow, and B. Ulanicki. Online Simplification of Water Distribution Network Models. In D. A. Savic, Z. S. Kapelan, and D. Butler, editors, *11th International Conference on Computing and Control for the Water Industry*, pages 749–754, Exeter, 2011. Centre for Water Systems, University of Exeter. doi: 10.2166/hydro.2012.029. 2.1.4

P. Pandya and V. Gupta. Enhancing Analog to Digital Converter Resolution Using Oversampling Technique. *International Journal of Innovative Science and Modern Engineering*, 2(5):37–40, 2014. 4.5

R. Pérez. *Identifiability and calibration of water network models.* PhD thesis, Universitat de Catalunya, 2003. 2.4, 4.5

R. Pérez and G. Sanz. Optimal placement of metering devices for multiple purposes. In *11th International Conference on HydroInformatics*, New York, 2014. 1.4, 2.6

R. Pérez, V. Puig, J. Pascual, A. Peralta, E. Landeros, and L. Jordanas. Pressure sensor distribution for leak detection in Barcelona water distribution network. *Water Science & Technology: Water Supply*, 9(6):715, dec 2009. doi: 10.2166/ws.2009.372. 2.6, 6.2.2

R. Pérez, F. Nejjari, V. Puig, J. Quevedo, G. Sanz, M. Cugueró, and A. Peralta. Study of the isolability of leaks in a network depending on calibration of de-

mands. In *11th International Conference on Computing and Control for the Water Industry*, pages 455–460, Exeter, 2011a. 1.1, 1.4, 2.3

R. Pérez, V. Puig, J. Pascual, J. Quevedo, E. Landeros, and A. Peralta. Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Engineering Practice*, 19(10):1157–1167, oct 2011b. doi: 10.1016/j.conengprac.2011.06.004. 2.7

R. Pérez, G. Sanz, V. Puig, J. Quevedo, M. A. Cuguero Escofet, F. Nejjari, J. Meseguer, G. Cembrano, J. M. Mirats Tur, and R. Sarrate. Leak Localization in Water Networks: A Model-Based Methodology Using Pressure Sensors Applied to a Real Network in Barcelona [Applications of Control]. *IEEE Control Systems*, 34(4):24–36, aug 2014. doi: 10.1109/MCS.2014.2320336. 2.7, 6.1.2, 6.2.2, 6.2.3, 6.2.4, 6.3.4, 7.2

R. Pérez, G. Sanz, M.-À. Cugueró, J. Blesa, and J. Cugueró. Parameter Uncertainty Modelling in Water Distribution Network Models. *Procedia Engineering*, 119:583–592, 2015. doi: 10.1016/j.proeng.2015.08.911. 1.4, 7.2

R. Pilcher. Leak location and repair guidance notes and... the never ending war against leakage. In *Water Loss 2*, 2007. 2.7

O. Piller. *Modeling the behavior of a network - Hydraulic analysis and sampling procedures for parameter estimation*. PhD thesis, University of Bordeaux, Talence, France, 1995. 2.2, 2.2.1

O. Piller and B. Bremond. A stochastic model for peak period analysis of pipe networks. In *Water Resources Planning & Management conference EWRSA*, Roanoke, 2002. 2.3.1

O. Piller and M. Propato. Slow Transient Pressure Driven Modeling in Water Distribution Networks. In *Water Distribution Systems Analysis Symposium 2006*, pages 1–13, Reston, VA, 2006. American Society of Civil Engineers. doi: 10.1061/40941(247)38. 2.1

O. Piller, B. Bremond, and P. Morel. A spatial sampling procedure for physical diagnosis in a drinking water supply network. In D. A. Savic and G. A. Walters, editors, *Water Industry Systems: Modelling and Optimization Applications*, pages 309–316, Exceter, 1999. 2.3.2, 2.5, 2.6

O. Piller, J. Deuerlein, D. Gilbert, and J.-M. Weber. Installing Fixed Sensors for Double Calibration and Early-warning Detection Purposes. *Procedia Engineering*, 119:564–572, 2015. doi: 10.1016/j.proeng.2015.08.909. 2.6

R. Pinzinger, J. Deuerlein, A. Wolters, and A. Simpson. Alternative approaches for solving the sensor placement problem in large networks. In *Water Distribution Systems Analysis 2011*, 2011. 2.6

R. Pudar and J. Liggett. Leaks in Pipe Networks. *Journal of Hydraulic Engineering*, 118(7):1031–1046, jul 1992. doi: 10.1061/(ASCE)0733-9429(1992)118:7(1031). 2.3.2

R. Puust, Z. Kapelan, D. a. Savic, and T. Koppel. A review of methods for leakage management in pipe networks. *Urban Water Journal*, 7(1):25–45, feb 2010. doi: 10.1080/15730621003610878. 2.7

C. Rahal, M. Sterling, and B. Coulbeck. Parameter tuning for simulation models of water distribution networks. *ICE Proceedings*, 69(3):751–762, jan 1980. doi: 10.1680/iicep.1980.2375. 2.3.2

Z. Rao and F. Alvarruiz. Use of an artificial neural network to capture the domain knowledge of a conventional hydraulic simulation model. *Journal of Hydroinformatics*, 9(1):15, jan 2007. doi: 10.2166/hydro.2006.014. 2.1.4

P. Reddy, K. Sridharan, and P. Rao. WLS Method for Parameter Estimation in Water Distribution Networks. *Journal of Water Resources Planning and Management*, 122(3):157, 1996. doi: 10.1061/(ASCE)0733-9496(1996)122:3(157). 2.3.2, 2.5

M. Romano, Z. Kapelan, and D. A. Savić. Geostatistical techniques for approximate location of pipe burst events in water distribution systems. *Journal of Hydroinformatics*, 15:634–651, 2013. doi: 10.2166/hydro.2013.094. 2.7

M. Romano, Z. Kapelan, and D. A. Savić. Automated Detection of Pipe Bursts and Other Events in Water Distribution Systems. *Journal of Water Resources Planning and Management*, 140(4):457–467, apr 2014. doi: 10.1061/(ASCE)WR.1943-5452.0000339. 2.7

L. Rossman. *EPANET 2 Users Manual*. Water Supply and Water Resources Division, National Risk Management Research Laboratory, 2000. 2.1, 2.1.1, 6.2.3

Safege. Piccolo Web Page. URL http://www.gfi.fr/utilities/PICCOLO.php. 2.1.3

J. Saldarriaga, S. Ochoa, D. Rodriguez, and J. Arbeláez. Water Distribution Network Skeletonization Using the Resilience Concept. In *Water Distribution Systems Analysis 2008*, pages 1–13, Reston, VA, mar 2009. American Society of Civil Engineers. doi: 10.1061/41024(340)74. 2.1.4

G. Sanz and R. Pérez. Aislamiento de fugas en una red de distribución de agua en función de la calibración de las demandas. *Tecnologia del Agua*, 32(341): 26–32, 2012. 1.1

G. Sanz and R. Pérez. Software webpage, 2013a. URL http://cs2ac.upc.edu/en/training-benchmarks/training-benchmarks. 1.4, 2.1.3, 7.2

G. Sanz and R. Pérez. Benchmark de control y supervisión de redes de distribución de agua. In *Actas de las XXXIV Jornadas de Automática*, pages 1–6, Terrassa, 2013b. 1.4, 2.1.3

G. Sanz and R. Pérez. Demand Pattern Calibration in Water Distribution Networks. *Procedia Engineering*, 70:1495–1504, 2014a. doi: 10.1016/j.proeng.2014.02.164. 1.4, 3.3

G. Sanz and R. Pérez. Parameterization and Sampling Design for Water Networks Demand Calibration using the Singular Value Decomposition: Application to a Real Network. In *11th International Conference on HydroInformatics*, New York, 2014b. 1.4

G. Sanz and R. Pérez. Comparison of Demand Pattern Calibration in Water Distribution Network with Geographic and Non-Geographic Parameterization. In *11th International Conference on HydroInformatics*, New York, 2014c. 1.4, 3.2

G. Sanz and R. Pérez. Sensitivity Analysis for Sampling Design and Demand Calibration in Water Distribution Networks Using the Singular Value Decomposition. *Journal of Water Resources Planning and Management*, page 04015020, 2015a. doi: 10.1061/(ASCE)WR.1943-5452.0000535. 1.4, 3.3, 7.2

G. Sanz and R. Pérez. Comparison of demand calibration in water distribution networks using pressure and flow sensors. *Procedia Engineering*, 119:771–780, 2015b. doi: 10.1016/j.proeng.2015.08.933. 1.4, 3.4, 7.2

G. Sanz and R. Pérez. Online Demand Calibration in Real Water Networks. *IEEE Control Systems*, 2015-SUBMITTED. 1.4

G. Sanz, R. Pérez, Z. Kapelan, and D. Savic. Leak Detection and Localization through Demand Components Calibration. *Journal of Water Resources Planning and Management*, page 04015057, sep 2015. doi: 10.1061/(ASCE)WR.1943-5452.0000592. 1.4, 4.6

D. Savic and G. Walters. Genetic Algorithm Techniques for Calibrating Network Models. Technical report, Centre for Systems and Control Engineering, Exeter, 1995. 2.3.2

D. Savic, Z. Kapelan, and P. Jonkergouw. Quo vadis water distribution model calibration? *Urban Water Journal*, 6(1):3–22, mar 2009. doi: 10.1080/15730620802613380. 2.3, 2.3.2, 2.3.2, 2.5.1

U. Shamir. Optimal design and operation of water distribution systems. *Water Resources Research1*, 10(1):27–36, 1974. 2.3.2

U. Shamir and C. Howard. Engineering Analysis of Water-Distribution Systems. *Journal of the American Water Works Association*, 69(9):510–514, 1977. 2.1.4, 2.3

M. Shimada. State-Space Analysis and Control of Slow Transients in Pipes. *Journal of Hydraulic Engineering*, 118(9):1287–1304, 1992. doi: 10.1061/(ASCE)0733-9429(1992)118:9(1287). 2.1

H. Sorenson. *Parameter estimation: Principles and problems*. New York, m. dekker edition, 1980. 2.4

D. Sumer and K. Lansey. WDS calibration and assessment for alternative modelling objectives. *Urban Water Journal*, 6(4):265–277, oct 2009a. doi: 10.1080/15730620802600932. 2.3

D. Sumer and K. Lansey. Effect of Uncertainty on Water Distribution System Model Design Decisions. *Journal of Water Resources Planning and Management*, 135(1):38–47, jan 2009b. doi: 10.1061/(ASCE)0733-9496(2009)135:1(38). 2.5

P. Swamee and A. Sharma. Decomposition of Large Water Distribution Systems. *Journal of Environmental Engineering*, 116(2):269–283, mar 1990. doi: 10.1061/(ASCE)0733-9372(1990)116:2(269). 2.1.4

K. Tang, B. Karney, M. F. Pendlebury, and F. Zhang. Inverse transient calibration of water distribution systems using genetic algorithms. In D. Savic and G. Walters, editors, *Water Industry Systems: Modelling and Optimization Applications1*, pages 317–326, Baldock, 1999. Research Studies Press. 2.3.2

E. Todini. Using a Kalman filter approach for looped water distribution network calibration. In D. Savic and G. Walters, editors, *Water Industry Systems: Modelling and Optimization Applications*, pages 327–336, Baldock, 1999. Research Studies Press. 2.3.2

E. Todini and S. Pilati. A gradient algorithm for the analysis of pipe networks. In *Computer applications in water supply*, chapter 1, pages 1–20. London, john wiley edition, 1988. 2.1.3

T. Tucciarelli, A. Criminisi, and D. Termini. Leak Analysis in Pipeline Systems by Means of Optimal Valve Regulation. *Journal of Hydraulic Engineering*, 125 (3):277–285, mar 1999. doi: 10.1061/(ASCE)0733-9429(1999)125:3(277). 2.3.2

R. Uhrhammer. Analysis of Small Seismographic Station Networks. *Bulletin of the Seismological Society of America*, 70(4):1369–1379, 1980. 4.1

B. Ulanicki, A. Zehnpfund, and F. Martínez. Simplification of Water Network Models. In *2nd International Conference on Hydroinformatics*, pages 493–500, Zurich, 1996. 2.1.4

J. Vítkovský, A. Simpson, and M. Lambert. Leak Detection and Calibration Using Transients and Genetic Algorithms. *Journal of Water Resources Planning and Management*, 126(4):262–265, jul 2000. doi: 10.1061/(ASCE)0733-9496(2000) 126:4(262). 2.1, 2.3.2, 2.7

T. Walski. Technique for Calibrating Network Models. *Journal of Water Resources Planning and Management*, 109(4):360, 1983. doi: 10.1061/(ASCE) 0733-9496(1983)109:4(360). 2.3, 2.3.2, 2.6

T. Walski. Assuring Accurate Model Calibration. *Journal of the American Water Works Association*, 77(12):38–41, 1985. 2.3

T. Walski. Case Study: Pipe Network Model Calibration Issues. *Journal of Water Resources Planning and Management*, 112(2):238–249, 1986. doi: 10. 1061/(ASCE)0733-9496(1986)112:2(238). 2.3.2

T. Walski. Standards for model calibration. In *American Water Works Association Computer Conference*, Norfolk, 1995. 2.3

T. Walski. Model calibration data: the good, the bad, and the useless. *Journal of the American Water Works Association*, 92(1):94–99, 2000. 2.3

T. Walski, D. Chase, D. Savic, W. Grayman, S. Beckwith, and E. Koelle. *Advanced Water Distribution Modeling and Management*. Haestad Press, 2003. 2.1, 2.1.4, 2.3.1, A.1, A.2

T. Walski, P. Sage, and Z. Wu. What Does it Take to Make Automated Calibration Find Closed Valves and Leaks? *World Environmental and Water Resources Congress 2014*, pages 555–565, 2014. doi: 10.1061/9780784413548.059. 2.3, 2.7, 5.2.3, 6.3.8

E. Walter and L. Pronzato. On the identifiability and distinguishability of nonlinear parametric models. *Mathematics and Computers in Simulation*, 42(2-3): 125–134, oct 1996. doi: 10.1016/0378-4754(95)00123-9. 2.4

G. Walters, D. Savic, M. Morley, W. Schaetzen, and R. Atkinson. Calibration of Water Distribution Network Models using Genetic Algorithms. *Transactions on Ecology and the Environment*, 19, 1998. doi: 10.2495/HY980131. 2.3.2

A. Wasantha Lal. Calibration of Riverbed Roughness. *Journal of Hydraulic Engineering*, 121(9):664–671, sep 1995a. doi: 10.1061/(ASCE)0733-9429(1995)121:9(664). 3.1

A. M. Wasantha Lal. Calibration of Riverbed Roughness. *Journal of Hydraulic Engineering*, 121(9):664–671, sep 1995b. doi: 10.1061/(ASCE)0733-9429(1995)121:9(664). 4.1

R. Wiggins. The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure. *Reviews of Geophysics*, 10(1):251–285, 1972. doi: 10.1029/RG010i001p00251. 3.1, 3.3, 3.3, 3.4, 4.1

Z. Wu and P. Sage. Water Loss Detection Via Genetic Algorithm Optimization-Based Model Calibration. In *ASCE 8th International Symposium on Water Distribution System Analysis*, Cincinnati, 2006. 2.7

Z. Y. Wu and Y. Song. Optimization Model for Identifying Unknown Valve Statuses and Settings. In *Water Distribution Systems Analysis 2012*, Adelaide, 2012. 2.7

Z. Y. Wu, P. Sage, and D. Turtle. Pressure-Dependent Leak Detection Model and Its Application to a District Water System, 2010. 2.7

W. Yeh. Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem. *Water Resources Research*, 22(2):95–108, 1986. doi: 10.1029/WR022i002p00095. 2.2, 2.4, 2.5

G. Yu and R. Powell. Optimal design of meter placement in water distribution systems. *International Journal of Systems Science*, 25(12):2155–2166, dec 1994. doi: 10.1080/00207729408949342. 2.6